IMPROVEMENTS TO A NEUROSCIENCE GRADUATE PROGRAM DERIVED FROM AN ANALYSIS OF PREVIOUS STUDIES OF QUALITY IN GRADUATE EDUCATION

1998

BEARDALL

UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799

APPROVAL SHEET

Title of Thesis: "Improving Graduate Education in Neuroscience"

Name of Candidate: Andrew Beardall
Neuroscience Program
Master of Science
1998

Thesis and Abstract Approved:

Jerome E. Singer

Jerome Singer, Ph.D.
Department of Medical & Clinical Psychology
Thesis Advisor

24 May 1998
Date

Brian Cox, Ph.D.
Department of Pharmacology
Committee Member

15 June 1998
Date

Donna Waechter, Ph.D.
Associate Dean, Medical Education
Committee Member

8 June 98
Date

Copyright Statement

The author hereby certifies that the use of any copyrighted material in the thesis entitled "Improvements to a Neuroscience Graduate Program Derived from an Analysis of Previous Studies of Quality in Graduate Education" beyond brief excerpts is with the permission of the copyright owner, and will save and hold harmless the Uniformed Services University of the Health Sciences from any damage which may arise from such copyright violation.

Andrew M Beadell

Abstract

Title of Thesis: Improvements to a Neuroscience Graduate Program Derived from an Analysis of Previous Studies of Quality in Graduate Education

Andrew McKenzie Beardall, Masters of Science, 1998

Thesis directed by: Dr. Jerome Singer, Department of Neuroscience

The question of how to determine the quality of an educational program has been examined for many years by everyone from scientists to popular magazines. In order to enhance the quality of a Neuroscience graduate program, a set of parameters to define "quality"must first be established. Drawing upon material from several scientific studies done over the last 75 years, this definition is described. Using those guidelines a series of recommendations is made as to how a university can increase the quality of its Neuroscience Program.

IMPROVEMENTS TO A NEUROSCIENCE GRADUATE PROGRAM

DERIVED FROM AN ANALYSIS OF

PREVIOUS STUDIES OF QUALITY IN GRADUATE EDUCATION

by
Andrew McKenzie Beardall

Thesis submitted to the Faculty of the
Department of Neuroscience Graduate Program of the
Uniformed Services University of the Health
Sciences in partial fulfillment of the
requirements for the degree of
Master of Science 1998

Table of Contents

Introduction

The job of the university is to provide a place for graduate education to occur. It provides space, resources, and money to support all of the programs which make it up. It is also responsible for services that are common to many parts of the school and would be a burden if each program had to provide for themselves. The university is also better equipped to make large investments that will serve the entire community. Yet there are still improvements that the university can make to better serve neuroscience.

A Neuroscience program is a complex entity in a graduate school. It is responsible for integrating individuals with a wide variety of experiences into a cohesive whole capable of advancing the frontier of knowledge in the field and educating a new generation of researchers, scholars, and teachers. Its parts come from many departments, anatomy, physiology, psychology, pharmacology, molecular biology, and biochemistry. The faculty are often active in their own field, but unite under the banner of "Neuroscience"so as to explore the possibilities offered by the nerves, the brain, and thought. A neuroscientist can specialize in the coding of DNA, the behavior of intracellular metabolism, interactions between nerves and other cells, the processes of brain activity, the structure of the neural system's components, perception, even thought and behavior. This vast array of possibilities can leave a graduate student overwhelmed and confounded. The program can help sort this confusion out and provide the best

possible education to that student.

Defining the best possible education is the hard part. An interdisciplinary program like neuroscience has difficulties that a program in a single discipline would not. The question of how best to serve the student and how to provide the best quality graduate training is of great importance to program directors and graduate deans. But in order to strive for high quality, the concept of quality must first be defined. Once a working definition of quality is established, the program can set goals consistent with that ideal. Improvements can be made at the level of the faculty, the program, and the university so as to best serve the student in a manner that will promote intellectual growth and expand knowledge about neuroscience, yet still stay within the available budget of time and resources. This paper will examine some of the major studies on rating the quality of graduate programs. From those studies indices of quality will be explored and suggestions on how a neuroscience program can increase its quality rating will be presented.

Before any attempt to define improvements can be made, the quality of a graduate program must first be ascertained. A program must have an awareness of its quality in relation to similar programs at other institutions, as well as an awareness of its own structure and policies. The more information that is available to the graduate deans and program directors, the better the decisions made on how to improve the training environment for the student.

The comparison of programs in different schools is an extremely difficult task in any discipline. To design a study that will be able to rate a graduate program in a fair and accurate manner is a Herculean task. It has not yet been done in a way that is acceptable to everyone in the graduate community and probably never will be. Creating a study that attempts to address all of the variables that can promote a successful graduate experience is expensive in terms of time, effort, resources, and money. A common criticism of such studies is that they fail to capture the "spirit" of a graduate program, the studies do not measure the interactive effects of the many variables. In essence, a study does not allow for the possibility that the sum is greater than the parts. Beyond the very real logistical difficulties of such a study there are the far more ethereal factors that can impede such an investigation. Issues of politics and ego can potentially confound research, and a study that actually compares or rates institutions is inviting the criticisms of the schools that do not rate as highly, no matter how fairly the process was designed.

Another difficulty with the comparison of graduate programs is the wide diversity that exists between programs that grant the same degree under the same name. A doctorate of Civil Engineering at the University A may focus on transportation planning or construction, whereas the doctorate of Civil engineering at the University B stresses materials or environmental engineering. A program can be influenced by the research interests of the faculty, the local business environment, or the source of funding. This diversity exists in many of the biological sciences especially in a program as interdisciplinary as neuroscience.

The greatest difficulty of creating a successful program is to create the goals to lead to that success(Rosovsky, 1990). Yet the process of designing these goals is often more difficult than implementing them. How does a department merge the many experiences and fields of a faculty into a neuroscience program? Once that task is accomplished, how is the department to know if they have a "good" program? There is no absolute measurement for the success of a program, only measurements that can be subjectively weighted in comparison with other programs.

Despite these difficulties, a number of investigators have endeavored to create a set of parameters that define quality. These parameters provide a standard that programs can measure themselves against in an attempt to enhance quality. The measures within the studies provide a guide to how a neuroscience program can improve itself on the level of the university, the faculty, the student, and the program.

Early Endeavors (1924-1957)

Some of the early attempts to compare graduate education programs were made by Raymond Hughes in 1924 and again in 1934. Hughes was the President of Miami University of Ohio, and in an effort to improve the guidance given to students about their potential graduate education he endeavored to evaluate the quality of education at 38 universities that offered the doctoral degree. Questionnaires were sent to distinguished scholars in twenty fields of study and the results were presented at the 1925 annual meeting of the Association of American Colleges (Hughes, 1934). This study was the first national ranking of graduate programs that had been published. The interest of the academic community was engaged by this work and there was quite a bit of criticism. Subsequent studies have also drawn a great deal of interest, but much criticism, sometimes even hostility, as well. However, Hughes, having been the first to attempt such a comparison, was greeted with curiosity by much of the scientific and academic community. The study's impact was extended to the general student population of the United States in 1928 when it was reprinted in the American Council on Education's first edition of *American Universities and Colleges* (Hughes, 1928). Presumably, a large number of college students had their graduate education plans influenced by this report.

In the 1800's, an American hoping to continue his education beyond the college level almost invariably had to go overseas for quality graduate work. These students brought back ideas such as the "scientific method," and tried to continue their work in

American Universities (Keniston, 1959). It was difficult at first to establish higher level learning programs at these colleges, due to obsolete equipment or insufficient libraries. However, in the late 19th century, Yale began offering advanced degrees, followed shortly thereafter by Harvard, Michigan, and Columbia. In the early years of the 20th century, this trend continued with the rapid expansion of American graduate institutions. Being the president of a university, Hughes wanted to know where on the continuum of quality schools fared in relation to one another. He had the faculty of twenty different fields at Miami create a list of schools that could be doing doctoral level work.[1] Next, a list of forty to sixty individuals who were teaching the subject in both colleges and universities was generated by the same faculty members. Letters were sent to these scholars with instructions to rank schools that taught graduate level education from a list of universities. There were varying degrees of success in returns ranging from a low of 16 in Geography to a high of 45 in Psychology. The returns were then weighed by rank on a four-point scale, much as a grade point average is calculated by today's student. The results were presented as a list of ranked schools in each discipline.

Though this work was groundbreaking, there were some clear problems with Hughes' methodology. He addresses some of these in his report. He concedes that the study is not exactly correct and should not be accepted as a universal truth, but that it is the educated opinion of about twenty to thirty individuals who are very well versed in the

---

[1]The methodology from Hughes 1924 study is drawn from Hughes' 1928 work included in *American Colleges and Graduate Schools.*

subject matter (Hughes, 1928). Some of the responding faculty also stated that the study would be far more accurate if large fields were further broken down into their component parts, such as breaking chemistry into inorganic, organic, bio and physical chemistry. Hughes agreed that this should be the subject of a more comprehensive study. The employment affiliation of the respondents was also a criticism from some of the faculty. Hughes dismissed this contention; however, he did nothing to control for this factor. Beyond the difficulties that Hughes addressed, there were many other criticisms. His study used small panels of distinguished scholars that were selected by his own faculty at Miami. The personal biases of his own faculty could easily affect who was considered distinguished. They were relatively diverse in terms of specialties within fields and types of universities, but most were trained and employed in schools in the Northeastern United States (Hughes, 1928).Of course, this observation is colored by the fact that most of the older and distinguished universities that offered graduate degrees in 1925 were located in the Northeast and Midwest. These criticisms notwithstanding, Hughes broke new ground by comparing graduate programs and schools across a large number of different disciplines.

Hughes also took the opportunity in writing this report to share some of his views of how to improve graduate education. He feels that the first improvement that could be made was that graduate schools should "make an effort to cultivate and strengthen the religious life of the graduate student."(Hughes, 1928). Graduate school of the early 20th century seemed to discourage the religious point of view in addressing life. Hughes

conviction that a moral and religious component was a necessary part of an educator is evident when he writes, "I am thoroughly convinced that no man should teach in an American college who is not sympathetic with religion and who has not developed his own religious ideals to such a degree that his life is guided by them."(Hughes, 1928).

The second improvement centered around specialization in the fields of the graduating students. Colleges of the time were looking for faculty members who possessed enthusiasm for the wide field in which they were educated, as well as an interest in related fields. Yet many of the new doctorates were only interested in the highly specialized area in which they have done their own work. Hughes suggests that graduate students who do not show a passion for research could focus their efforts upon gaining a comprehensive knowledge of many aspects of their field, as opposed to mastery of a single specific domain. They could receive a different degree that reflects this course of study and according to the President of Miami of Ohio University, ". . . such men would find a ready welcome on the staffs of American colleges."(Hughes, 1928).

Hughes' third improvement focused on the contribution that graduate schools made toward preparing their students as teachers in an educational environment. Since a considerable percentage of students in graduate school find employment in academia, it is logical that graduate schools should teach some of the basics of education, such as how to share information in an interesting way or how to plan an effective curriculum. "None of us as college presidents, or deans, or professors desire to add to our staffs men who are

not capable, enthusiastic teachers; nor . . . who are not interested in students as human beings"(Hughes, 1928). These words carry great weight coming from a man of Hughes' position, in fact, it is a wonder that colleges do not demand that graduate schools better prepare their students for a role in the education process. It is ironic that the aspect of a professor that will directly affect the most people has so little time devoted to it in the graduate education process.

Finally, Hughes feels that all graduate students should be instilled with a love for learning toward all disciplines, a broad interest in knowledge as a whole, and a modicum of social training. The cultured graduate student should be a role model for incoming undergraduate students, for the ". . . crude youths who enter our doors" (Hughes, 1928). He believes those living accommodations for graduate students should be available on the university campus to allow them to interact with individuals from other disciplines and support one another in the graduate community. These shared experiences would give the student additional perspective on his own field and help him to achieve the goal of learning to think.

Hughes' opinions are all thought out and for the times, appropriate; but how have they translated into today's scientific education? Religion is not a part of graduate education today, though ethics is often a required part of the curriculum. A broad knowledge of a field is often encouraged in the early phases of a program, with courses and rotations in different aspects of a field as required components and later in the

program with continuing seminar presentations. Neuroscience, in particular, must accomplish this goal since the field itself is so diverse. Creating a graduate committee is done at many universities by providing housing, graduate student social events, and research seminars in which students can share their work with those outside their field.

However, Hughes' third point, regarding education in education' has not been so well addressed. "I still feel . . . that not a few [graduates] are coming somewhat imbued with the idea that students are a nuisance and interfere with work . . . "(Hughes, 1928). Many students today would agree with this assessment of their teachers. Hughes feels that there was more to teaching a subject than simply understanding it. Aside from being informative, a teacher must be interesting; as much as it is the student's responsibility to want to learn, a teacher must create an environment that is conducive to learning. Had more graduate schools listened to Hughes' recommendation on this topic, students today would not still be chanting the mantra of, "My professor is really smart and knows the material, he just cannot teach it." This is not to say that there are no effective teachers in the college environment. There are many wonderful, enthusiastic teachers who can pass on their knowledge in a compelling, and sometimes even fun, manner. However, if asked, the average undergraduate will find this scholar the exception rather than the rule. That is not to say that others are incapable of this feat, but that they may just need some training on how to accomplish it.

The American Council on Education's Committee on Graduate Education

commissioned Hughes to do a follow-up study in 1934. Hughes tried to address many of the criticisms of his earlier study in this subsequent venture.[2] The number of fields examined expanded from 20 to 35, tripling the number of Biological Sciences examined. Every school that was known to be offering graduate level work, based on the reports of graduate deans, was included in the survey. One hundred respondents were selected for each field based on a list provided by the secretary of a national professional society (There were a few exceptions for smaller disciplines, for example Aeronautical Engineering had only 49 surveys sent out). As opposed to ranking the programs, this study asked the respondents to indicate whether the department provided an adequate graduate education in the field. They were also given the option of highlighting any programs that were especially distinguished in the field. The classifications were determined by a simple majority, if a program were determined distinguished by more than 50% of the responses then it was given a distinguished rating. If a program was determined adequate, but did not qualify as distinguished, by more than 50% of the responses then it was given an adequate rating.

Just as in the earlier study, Hughes and his committee had some thoughts to share about the future of graduate education. A major focus in this report was the differentiation between a graduate school and a professional school. They felt there was a distinction between an advanced professional school, such as in Medical School, in which

---

[2]The methodology from Hughes 1934 study is drawn from Hughes' 1928 work "Report of the Committee on Graduate Instruction."

the goal is the clinical degree, and a graduate school in a science, in which the primary goal is the promotion of research. Though both types of programs can exist at the same institution, they are often autonomously administered and the material covered is taught independently. Both programs could benefit from the input of the other, and the committee recommended that departments doing graduate work within professional schools should be treated as units within the graduate school. A graduate school should also maintain close relations with research institutions and vice versa, since both will be able to benefit from such an association. The institution of research can provide "Experience, apprenticeship, and broad training . . . "(Hughes, 1928) in areas that may be beyond the scope of the university's program or finances. The institution will benefit in that the faculty will see that their work can have an educational value. The graduate schools will benefit because they will produce doctorates with greater experience and an understanding of a professional research environment. Most of all, the student will benefit from a wider range of opportunities and the chance to make associations in the professional arena that will be helpful to him after he receives his degree.

The commentary by Hughes in this second study is quite different from the first. In the 1925 study, he focused on how graduate schools could improve themselves from the inside. Increasing student interaction, teaching how to educate, and the other suggestions that he made are all improvements that could be made by a program or a department internally. The ideas in the 1934 study look at more of the big picture, how a graduate program can beneficially interact with entities that are not a part of the graduate

school but have similar interest in the field of study. A professional school and a graduate program could share resources if it was appropriate, exposing students of both types to the other. This exposure will show the students that there are other applications to the field in which they study. If a graduate school and an institution of research associate, the students will have a wider range of research options available. Hughes always seems to have the student in mind when making suggestions, for he believes the role of a graduate school is to produce the best doctorates possible, talented on a social and educational level, as well as in research.

Hayward Keniston, of the University of Pennsylvania, took the next step in 1957. The University of Pennsylvania was among the 25 institutions in the Association of American Universities and Keniston wanted to find out the relative position of his school in this organization[3]. As opposed to Hughes' study, in which a group of distinguished scholars was consulted, Keniston chose to poll the chairmen of 24 different departments at each university. He selected chairmen because it is part of their job to be familiar with programs at other universities. A chairman needs to be well informed, so that his responsibilities in recruiting faculty and advising students can be properly executed. Keniston asked each of the chairmen to rate the strongest departments in his respective discipline, either numerically or in groups of the first, second and third set of five out of the total of 25. About 80% of those sent the survey responded. The scores were assigned

[3]The methodology from Keniston's study is drawn from Keniston's 1959 work, *Graduate Studies in the Arts and Sciences at the University of Pennsylvania.*

by rank if the response had been in numerical ranks, or by an average rank score if the response had been in groups of five. Individual programs were combined to define scores in the subject areas of Biological Sciences, Humanities, Social Sciences, and Physical Sciences. These subject scores were combined to create overall university graduate program scores.

As with the previous studies, the results were very much debated. There was wide criticism of this work centering around the use of the department chairmen as the source of data. In general, chairmen tended to be older, full professors who could exercise a great deal of control over hiring in their institution. As individuals who have been in the field for a long time, a time lag could be built into the sample. Institutions could be judged based on the quality that they have traditionally possessed as opposed to their current status. The small sample size also magnifies the effects of sample bias. The chairman's opinions tended to show favor to region, alma mater, and current institution (Hughes, 1928). Keniston concedes that the individual scores lack validity and that this study is a more valuable tool for studying trends, especially when compared to the Hughes studies of a generation before. The benefit of creating aggregate scores for the entire university was also questioned. The prospective student might like the prestige of going to a university which is highly rated, but it is the strength of the individual program that should mostly influence his decision.

Recent Studies (1963-1982)

In 1963, the Commission on Plans and Objectives was created by the American Council on Education. The purpose of this body was to study the long range policies of higher learning. One of the primary focuses was the strengths and weaknesses of current graduate institutions in providing future teachers and researchers. The answer to this question was published in 1964 in a book called *An Assessment of Quality in Graduate Education, a Comparative Study of Graduate Departments in 29 Academic Disciplines* published by the American Council on Education (ACE) and chaired by Allen Cartter.

Before beginning this study, Cartter set three goals to achieve. The first was to bring the earlier studies up to date. Programs change in seven years. Rankings should reflect this change. The second goal was to widen the scope of the study to include all of the major universities in the United States. The 38 schools studied by Hughes and the 25 by Keniston did not allow for the dozens of other institutions that could potentially have distinguished graduate programs. The third goal was to learn as much as possible about utilizing subjective measurements so as to improve data collection by these methods in the future (Cartter, 1964). Hughes and Keniston showed where some of the pitfalls of subjective measurement can be found and the 1963 study tries to examine these criticisms for any validity.

These are all worthy goals for Cartter to aspire to. However, is there a larger

reason to proceed with this study? Or will this work result in nothing more than a list that universities will use as an ego booster? Cartter himself states, "The diversity of the American system of higher education has properly been regarded by both the professional educator and the layman as a great source of strength, since it permits flexibility and adaptability and encourages competing solutions to common problems."(Keniston, 1959). A variety of approaches to examining a problem can often lead to a correct answer, yet the differing processes to achieve that answer can expose different aspects of the same issue. The wide range of skills, finances, and talent make each university a different place, and therefore each Ph.D. that is produced will be a reflection of those differences. Therefore, there will be a wide range of talents in graduates from programs at different institutions. Providing ratings of quality helps the institution and the prospective student. The university can benefit by learning its own strengths and weaknesses and seeing other schools that have surpassed shared problems. The student can use ratings to determine the best program for him. Ideally the student wants an educational, productive, happy, and even fun experience in graduate school. The educational opportunities, level of work, programs offered, research topics pursued, and financial aid available are all essential components that are factored into selecting an appropriate graduate program. For the student to select an institution that is not matched to him and for the university to accept him, does a disservice to them both and will not provide a great benefit to either. This study was created to help universities in the process of self examination and to provide a resource to students.

With the question of why to examine quality answered, the next step to address is how to measure the elusive concept of quality. There are very few measurements that are truly objective even if they appear to be on the surface. The University of Maryland at College Park's McKeldin (Graduate) Library has a 1.6 million volumes (Dopp, 1997). That is quite an impressive number, but is it a measurement of quality? The University of Maryland has been offering graduate degrees since 1920 (Dopp, 1997). and presumably has been amassing books in its library since then. Many could be out of date or the collection on a specific topic could be insufficient for graduate education. Though the number of volumes is an objective measurement, it is not necessarily an index of quality for a graduate level education. Similar reasoning can be applied to number of faculty, students, grants, awards, and publications. Though distinguished universities tend to have more of the above rather than less, numbers alone are not indicative of quality.

The faculty are the cornerstone of the graduate program upon which everything else is built. Their interests provide research opportunities, their research provides publications, their publications provide distinction, their distinction provides grant money, and money provides recognition for the entire university. A department can possess many notable faculty, even giants in their respective fields, who would bring prestige to the university and the department. Yet an individual measure of each faculty member does not reveal their interaction as a whole department. If a department has several luminaries who cannot work collectively, it will not provide nearly as good of an educational environment as a school with competent faculty who can work in unison. It

is also not fair to judge an entire group of faculty based on a few well-known members. It is clear that just looking at the faculty cannot provide a clear measure of quality either.

The question of how to measure quality cannot be answered with objective means, which forced Cartter to use subjective indices.[4] He then had to address how to do acquire subjective assessments as fairly as possible. One of the biggest changes from the previous studies was the nature of the sample pool. As opposed to a relative handful of top scholars or department chairmen, the 1964 study included responses from more than 4,000 individuals, including department chairmen, senior scholars, and junior scholars. Additional questions to assess background factors, such as alma mater, of each respondent were included in the questionnaire. "Objective" measures such as library resources and faculty income were also gathered. Finally, to take into account how quickly subjective ratings can change, a follow-up study was planned for five years later.

The participants were selected from a list of department chairmen, senior and junior scholars that was requested from the academic deans of the 106 graduate schools in the study. One hundred of the institutions that were participating were members of the Council of Graduate Schools in 1961; the remaining six universities had all granted at least 100 doctorates within the last decade. The selection of the number of scholars in the

---

[4]The methodology from Cartter's study is drawn from Cartter's 1964 work, *An Assessment of Quality in Graduate Education, a Comparative Study of Graduate Departments in 29 Academic Disciplines.*

sample from each university was weighed by the number of Ph.D.s that the school had awarded in the last decade. The three primary questions Cartter asked examined the subject's assessment of quality of the faculty, rated on a scale of 1 (distinguished) to 6 (not sufficient to provide acceptable doctoral training); the effectiveness of the institution's doctoral program, rated on a scale of 1 (extremely attractive) to 4 (not attractive); and the projected changes of the departments relative to one another in 5-10 years, rated on a scale of 1 (relative improvement) to 3 (relative decline). Each respondant also had the option of answering "insufficient information." The questions were followed by a worksheet-type list that included each of the universities that had awarded at least one doctorate, based on reports to the Office of Education, in the respondent's field of expertise on which the respondent entered his ratings. On April 1, 1964, 5,367 questionnaires were mailed out, with follow-up reminders sent out in mid-May and mid-June. The response was 80 percent with about 75 percent useable. Some were returned due to incorrect addresses or death of the potential respondent (one was even returned by a widow, explaining that her husband had been dead for over two years). A questionnaire was considered useable if the first two questions were answered.

The results were tabulated and each response was given a numerical weight, calculating a score for a university much as a student would calculate his grade point average. Once the program ratings had been calculated, they were divided into categories ranging from not sufficient to distinguished for the first question, and from not attractive to extremely attractive for the second question. Finally, the results were arranged in

tables by discipline, with "distinguished" and "strong" programs given a numerical ranking. The scores were further broken down by the type of subject, chairman vs. senior scholar vs. junior scholar.

Then came the difficult job of addressing several concerns about potential bias in the data. The authors went into extreme detail concerning how they checked the data in four of the fields, of which physics was the only science examined. Alma mater and current institution were a source of concern to the investigators, for a scholar might be tempted to rate his own school higher. When more than four respondents received their Ph.D.s from the same university their input was dropped, the ratings recalculated, and the impact of the alumni analyzed. A similar procedure was executed dropping the present employees of a university. The alumni rated their former schools an average of 17 percent higher than the main sample whereas the employees rated their current schools about one-third higher. Once a set of ratings was created, a set that excluded "personal attachments," there was little change from the raw ratings that included all of the scores. Of 86 departments, 57 did not change rank, and of the 29 remaining, none changed by more than two spots in rank (Cartter, 1964). Though affiliation did have a significant effect on the rating of individual universities, the overall effect was negligible due to the large sample size.

The next factor examined was the regional variation in the ratings. The study divided the country into four regions, East, South, Midwest, and West and the ratings

were compared by area of origin. The "Distinguished" Universities showed little variation, however the next level, the "strong" universities, had a considerable amount of dissent depending on the region. Raters from the East and Midwest tended to be relatively close to the overall averages, whereas those from the South tended to be far more generous and those from the West tended to be far more critical (Cartter, 1964). This trend was not the case in all of the disciplines examined. In the three non-sciences that were examined in detail, the West had been the most favorable in their ratings (Cartter, 1964). Respondents should be selected from a wide area, without a strong concentration in one region. This concentration could skew the results.

The final detailed examination was to investigate the impact of departmental prestige on the ratings. The "distinguished" departments were quite well known and were rated by at least 95% of the respondents (as opposed to "insufficient information"). However, there was a much wider variation in the lower ranked universities. Prestige confers a certain degree of publicity to a program. The programs less well known to the academic community were not as highly rated and received more "insufficient information" responses (Cartter, 1964). If a scholar lacks knowledge of a department or its faculty, it can be interpreted as less distinguished. This does not necessarily imply a lack of quality, just a lack of prestige.

After completion of analysis, several patterns developed among the more highly rated programs. One of the most obvious was how similar programs tended to reinforce

one another in the ratings. If a university had a "distinguished" rating in one field, the other fields that were closely related tended to be highly rated. For example, MIT's mechanical engineering department was rated as distinguished as was its electrical engineering program. The similarity in ratings of allied fields can be a reflection of a school's financial commitment to the overall discipline, shared faculty, or common resources.

Just as regional variation is a factor in rating the programs, it is also a distinction that can be seen looking at the results. About 80% of the "distinguished" programs are located in just five states. When expanded to include the "Strong" universities, every area of the United States had at least one program that fell into one of these categories. However, the single "strong" program in the Rocky Mountain area does not compare to the 215 "strong" and "distinguished" programs in the Northeast. The Southeast and Southwest are also lagging behind in highly rated programs in proportion to their population and number of universities (Cartter, 1964). Cartter also noted how the nation's capital had no "distinguished" or "strong" universities within its borders; in fact, none rated above "adequate plus."

Finances are another consideration of quality. Cartter's report showed a correlation between salary and rating of the institution. Highly-rated institutions seemed more willing to spend larger amounts on filling faculty positions; in some cases up to 50% more (Cartter, 1964). The cost that is associated with each student is a considerably

more murky question. In attempting to ascertain the amount spent on each student, factors that are not routinely measured by a university must be examined. What is the difference in resource allocation between graduate students and undergraduates? What is the difference in cost of laboratory-extensive programs, such as chemistry, versus non-laboratory-extensive programs such as classics? How much is a part-time student costing as opposed to a full-time student? These are questions that are often not asked, much less explored or described in financial detail. A far more detailed department-by-department study would have to be done to address these points.

The library resources are essential to the graduate student, but Cartter realized that size alone could not determine the quality of a library. He added two other factors that, when averaged with the total number of volumes, created a score called the "Overall Library Resource Index." The first additional factor was the number of volumes added per year. This score was a reflection of how current a library's holdings would be. The second addition centered on the number of periodicals to which the institution subscribed. This score helps to even the playing field for the science students for whom the journal is the primary library research instrument. As expected, universities with higher "Overall Library Resources Index" scores tended to rate higher. The notable exceptions to this pattern were the "Technology Institutes," such as MIT and Cal Tech, which often received high ratings but whose specialized library resources resulted in a lower overall library resources score (Cartter, 1964).

Finally, Cartter asked how an institution could improve its quality rating. He shares some of his speculations based on the collected data. Firstly, a university must be sufficiently well known for people to rate it (Cartter, 1964). A reputational study can only measure a school that has a reputation that others in the field recognize. Many schools were well known and well rated locally, yet lacked the national prominence to be highly rated universally. Secondly, there was a strong correlation between salary and quality (Cartter, 1964). Programs that have disproportionately high salaries for their rating are prime candidates for improvement in the next study. Thirdly, universities that have disproportionately high library scores for their quality ratings also have potential to quickly improve (Cartter, 1964). These views are just conjecture, but since a follow-up study was to be done within five years, there would be a basis for comparison.

Cartter's study expanded and improved upon earlier work by Keniston and Hughes. The number of schools was greatly increased and efforts were made to rectify "objective" measures with the subjective scores based on reputation. Some of the criticisms of earlier studies were addressed by expanding the pool of respondents not only in size but in academic rank so as to provide a wider array of responses. Potential Response biases such as affiliation and regional proximity were examined. Cartter tried to not only build upon earlier work and help define the quality of graduate education, but he set the stage for a follow-up study to build on his work.

The mantle was passed from Cartter to Kenneth Roose and Charles Andersen and

in 1969 they did a study that resulted in the publication of *A Rating of Graduate Programs* in 1970. In his foreword, Logan Wilson, President of the American Council on Education, reflects that it was gratifying that many universities used the information presented in the 1964 study to improve their own graduate programs. Various departments were able to use the report as a resource or reference when planning budgets, allocating existing funds, and creating long term plans. Further, more than 26,000 copies of Cartter's report were published, many of which found their way into the hands of students, whose decisions were influenced by the report.

The procedure for the 1969 study was very much the same as Cartter's.[5] Roose and Andersen utilized the identical criteria for program inclusion in the study. The university must have awarded at least 100 doctorates in the ten academic year period from Fall 1957 thru Spring 1967. These dates were selected because it was the most recent period for which the information was available from the National Research Council. Under these criteria, additional universities were eligible, increasing the number of schools from 106 to 130 and the number of programs from 1,663 to 2,229. Inclusion of seven additional fields in the humanities and biological sciences brought the final number of programs to 2,626. Inclusion of a program on the questionnaire required that the program have awarded at least one doctorate in the ten-year period previously described. The three new fields in the biological sciences had special criteria created for them since

[5]The methodology from Roose/Andersen's study is drawn from Cartter's 1970 work, *A Rating of Graduate Programs*.

no data was available for awarded doctorates. Schools were included if the institution had awarded doctorates in similar related fields in the ten-year period. For example, the field of developmental biology was included for schools that had awarded degrees in genetics or embryology.

Participants were selected by first gaining the assistance of the graduate deans. Of the 130 deans who were asked for input, all replied. Each was asked to submit the names of chairmen, senior scholars, and junior scholars, the number dependent on how many doctorates the institution had awarded in relation to the total. Each of the people submitted was mailed a questionnaire packet. The "objective" portion of the packet was identical to the Cartter version, and the rating portion had only one major change. The first two questions, concerning quality of faculty and effectiveness of the doctoral program, were the same; the variation was in the third question. In Cartter's study the third question asked for the respondent's forecast about the potential change in the school over the next 5-10 years. Roose and Andersen instead asked if there had been any change over the last five years. The respondents could answer on a scale of 1 (Better than 5 years ago) to 3 (Worse than 5 years ago), there was also a fourth option for "insufficient information." In the spring of 1969, 8100 questionnaires were mailed out in two waves. Follow-up mailings were done in late May/early June and in late July/early August. Analysis of results began in late August, after which no further responses were accepted. Of the 6325 questionnaires returned, slightly less than 6100 were useable.

Scores were weighted using the exact same method Cartter had used, as if calculating a grade point average. However, the scores were then rounded to the first decimal place instead of the second. It was Roose and Andersen's assertion that a few hundredths of a point is inconsequential and serves no purpose. In the analysis of the ratings they tried to de-emphasize the rank order relationship, seeing it simply as a tool for raising or lowering egos (Roose, 1970). They did keep ranks for the top schools so as to compare them to the Cartter report, but the remainder were simply listed in alphabetical order by category. Gone were the labels of "Distinguished," "Strong," "Good," "Adequate Plus," "Marginal," and "Not Sufficient" for the first question, replaced with numerical ranges. Institutions that scored 3.0 or higher corresponded with the highest two classes of the previous report and were given ranks. "Good" became "2.5-2.9" and "Adequate Plus" became "2.0-2.4". The remaining institutions were not listed. A similar system of numerical ranges was established for the three point scale used to answer the second question. Insufficient information responses were dropped and not included in the calculation. The third question was not scored, instead a distribution was created based on the percentage of each possible response.

The in-depth validity analysis that Cartter performed was not done in this study. The authors felt that its close similarity in both procedure and instrument made the additional expense unnecessary. However, it would have been interesting to see some further effort expended in comparing the programs featured in both studies. Though the 1969 report did show the change in ranking for the schools that were examined in both

studies, none of the "objective" data was presented. Did the schools with greater library resources or higher salaries improve in standing as Cartter predicted? What "objective" improvements (perhaps influenced by Cartter's report) have been made over the five years in between the studies? Many programs improved; it would have been helpful to other programs seeking to better themselves if some of the objective changes had been presented. In addition, a validity test should have been done on the third question, concerning the change in quality over the five years between the studies. Checking the objective data for any changes that could indicate a change in quality or consulting an expert panel from a professional society are two possibilities that were not included in the final report.

Nevertheless, the Roose-Andersen report was a well done follow-up to the Cartter study. It expanded the scope to include more fields and the inclusion of more schools and programs showed the growth of American graduate schools. Most importantly, it showed that the scores from Cartter's report could not be written in stone. Graduate programs were constantly changing and growing, many of them for the better. It must have been quite a reward for the faculty of a program, who after implementing changes based on the Cartter report, received a higher score five years later. Roose and Andersen's study results probably helped to validate the efforts of many a dean and scholar alike who wanted to improve their programs.

As useful as a follow-up report was proven to be, another was not done until

1982. Lyle Jones, Gardner Lindzey, and Porter Coggeshall took command of this endeavor under sponsorship of the American Council of Learned Societies, the American Council on Education, the National Research Council, and the Social Science Research Council. The result was a five-volume series called *An Assessment of Research-Doctorate Programs in the United States,* each report detailing a different field of graduate education. Though neuroscience was not yet a topic examined, six different fields in the biological sciences comprised the fourth report in the study. Of the 2669 programs in 32 disciplines that were examined, 616 fell under the topics of botany, biochemistry, cellular/molecular biology, microbiology, physiology, and zoology and were included in the study.[6]

In the previous two studies more fields in the biological sciences were examined than in this current one. Cartter used seven fields and Roose and Andersen expanded this number to ten. The 1982 report was limited to six fields in the biological sciences. The decision to limit the number of fields was a financial one, since the project had funding to focus on about thirty areas of study spread over the five divisions of graduate study. Once the designers of the study had been limited to six biological sciences fields, they had to determine which would be included. The decision was made based on the number of doctorates awarded in from 1976-1978. This data was available from both the Educational Testing Service (ETS) and the National Research Council (NRC). The

---

[6]The methodology for Jones/Lindzey/Coggeshall's is drawn from their 1982 work, *An Assessment of Research-Doctorate Programs in the United States*: Biological Sciences.

information from the latter was more detailed, in that there were more specialized categories into which a recently awarded Ph.D. could be classified, so the NRC data was the primary source used for determination of field selection for this study. Additional difficulties were caused by the trouble differentiating fields. What one individual would consider immunology another might consider cell biology, for example. Therefore, the investigators decided to use traditional delineations along departmental lines to determine the fields examined, as well as incorporating some programs from anatomy, biochemistry, biophysics, cell biology, developmental biology, and genetics into the cellular and molecular biology category.

In order to be included in the study a program had to award a specific number of doctorates between 1976 and 1978. Alternately a program could be included if it awarded a set number of doctorates in that field in 1979, usually about one-third of the 1976-1978 amount. These requisite numbers were calculated so that the programs that met the requirement would have awarded at least 90% of the doctorates in the field. The list of programs was then sent to the study coordinator, usually the dean, at the school. The coordinator was asked to make any modifications for programs that may have been overlooked or programs that were no longer active. Only 15 out of 243 schools contacted did not participate, eight because they did not return the materials on time and seven because they declined to be a part of the study. In previous studies, none of those programs had received "Strong" or "Distinguished" ratings, and a list of these institutions that met the criteria but did not participate was included in the results portion of the

report. From the individuals who did respond, the final list of 616 programs was constructed. The final decision of whether or not to include a program was left to the institutional coordinator and this led to some problems, especially in cellular and molecular biology, due to the diversity of programs that field can represent. This difficulty leads to the concern that some programs might have been left out.

The study measures each program in six different areas with detailed points leading to sixteen dimensions spread over the areas. The following general areas were examined: program size, characteristics of graduates, reputational survey results, university library size, research support, and publication records. This large number of characteristics is quite different from previous studies. Until this report the reputational standings were the crux of the rating for a program, using other factors to reinforce the conclusions drawn from them. Now, all of the areas were given substantial weight in the rating process. In addition, the 1982 report examines the student in a more thorough manner, believing that they are a reflection of program quality.

There were three dimensions under the general topic of program size. Each of these pieces of data was provided by the participating universities in the study. The first is the number of faculty members that held the rank of assistant, associate, or full professor who had a significant contribution to the graduate education process. The second is the number of Ph.Ds produced from 1975-1980. The third asks how many graduate students are currently enrolled in the program intending to earn a Ph.D. Detailed

instructions were included on how to complete this section, specifying guidelines as to what is meant by a "significant contribution" to graduate education as well as rules about inclusion of emeritus and visiting faculty and faculty on sabbatical. Previous studies had shown a high correlation between program size and reported quality and that topic is addressed in the analysis of this study.

Four dimensions under the topic of characteristics of graduates were examined in the study. The first is the fraction of students who received national fellowship or training grant support while a graduate student. The second was the median number of years that it took for students to earn their doctorates. The third was the fraction of students who at the time of graduation had made commitments to employment. The fourth was the fraction of graduates who had committed themselves to a job in a university setting. This information was gleaned from the NRC's Survey of Earned Doctorates and was the cause of much debate among the committee due to the lack of precision in the data. The question about fellowships is complicated by the fact that methods of supporting graduate students vary so widely from one university to the next. In a large public university there may be more teaching assistantships available, whereas in a smaller research university there could be more research assistantships available. Availability of one or both of these options could mean use of fewer federal training grants. The number of years to achieve the Ph.D. is used under the mindset that the most talented students will finish the degree more quickly. This frame of reference is fine for comparing students within a program, but in comparing students between programs it

would not be a good measure due to the simple fact that programs are so different. A physiology program at University A could have more stringent requirements for the degree or lousy advising and student support program when compared to University B. Either of these factors could add a considerable amount of time to the duration from enrollment to Ph.D. at University A. The third and fourth measures both proceed from the assumption that the most talented students will find jobs the most quickly. The fourth measure specifically tries to ascertain the number of students going down the traditional path to becoming researchers in academia. Naturally, not all research is done in academic settings and the authors are careful to note that other lab settings in government and industry provide fertile ground for becoming successful in research. Both measures are also confounded by factors such as individual job preference, job availability, and financial need. However, all of these measures do try to examine the final product of the graduate program--the student--as an index of the quality of a program. The idea that the student can be used as a measurement of quality is a very good one, but in this case the measurements needed to be better refined.

Three of the dimensions in the reputational survey should be familiar by now. Quality of the faculty, effectiveness in education, and improvement in quality over the last five years are all questions that could have come straight from Cartter's or Roose and Andersen's studies. The investigators added in a fourth measure, questioning the familiarity of the respondent with the program that he was being asked to rate.

The reputational survey was mailed out in April of 1981 to 1848 faculty members at 228 universities. This number represented about 15% of the total number of faculty in biological sciences at all of the schools in the study. The sample was selected based on the number of faculty members and the number of doctorates produced by that university and at least one evaluator was selected from each university. The number of respondents of each faculty rank was selected to remain consistant with the proportion of each rank in the total number of faculty.

The survey instrument consisted of a questionnaire that listed a random sample of programs in the respondent's field of expertise. There were fifty programs listed on each survey, and each program was included in 150 of the questionnaires mailed. There were a few other changes in the survey besides the addition of the question about the familiarity of the respondent with the program. A question asking the evaluator's field of specialization was added to allow comparison between program evaluations from individuals in different specialties in the same general discipline. However, the most significant difference in the questionnaires in this study was that the respondents were provided with information on the programs that they were evaluating. This information consisted of the number of doctorates that were produced by that program over the last five years, as well as the names of the faculty members. This addition was made on 90% of the instruments mailed out, leaving 10% without the extra information for comparative purposes. The final modification to the questionnaire was the exclusion of the evaluator's own program from the survey that was received. This was done to eliminate

the criticism about bias toward the university in which the respondent was currently employed.

The questions asked of the respondents were very similar to previous studies. In fact the first two questions were almost identical to Cartter's survey. The first examined the subject's assessment of quality of the faculty, rated on a scale of 0 (Distinguished) to 5 (Not Sufficient to Provide Acceptable Doctoral Training). The second examined the effectiveness of the institution's doctoral program, rated on a scale of 0 (Extremely Attractive) to 3 (Not Attractive). These questions differ from Cartter's only in that the possible ratings were from 1 to 6 on the first question and 1 to 4 on the second (The number assigned to each rating was shifted by one place, but the total number of possible ratings remained the same). The third question more resembles the Roose and Andersen question examining the improvement over the last five years. The evaluators are asked to rate improvement in both the quality of the faculty and the effectiveness of the program on a scale of 0 (Poorer than five years ago) to 2 (Better than five years ago), with a provision for not enough information. This question could have been more informative had it been split into two questions examining the improvement in the aspects of faculty quality and training effectiveness individually as opposed to collectively. The final question asked the respondents to rate their familiarity with an institution on a scale of 0 (Little or No Familiarity) to 2 (Considerable Familiarity). This new addition provides an additional factor in the analysis and weighing of responses from the evaluators.

The response rate was considerably lower than previous studies. Only 56%, down approximately 15%, returned completed survey questionnaires, with a total of 1026 useable responses. The authors postulate two reasons that the response rate was much lower than previous studies. The first is derived from the ultimate limitation on any experiment, expense. Since each survey was between 20 and 30 pages, duplicate copies were not sent to evaluators who did not return the first survey in a timely manner. Though a follow-up letter was sent, the cost of printing and mailing a duplicate survey was prohibitive and was not done unless specifically requested by a respondent. Perhaps without receiving a second notice, the respondents did not feel the sense of urgency or guilt that would have encouraged them to respond. The second factor that lowered the response rates was the overall feeling of the educational establishment regarding reputational measures. The dissatisfaction with this data collection technique was the primary reason that a multidimensional approach was used in this assessment. Had the survey been designed in a different manner, perhaps stating that the reputational assessment was but one of a series of measures, the response rate might have been higher. There was no difference in response rates between the different ranks of faculty nor the "informed" versus the "uninformed" survey forms.

This portion of the study shares many of the same criticisms that previous reputational studies have incurred. The designers attempted to circumvent some of the problems in the design of the study. The evaluators were not given their own university on the survey so as to eliminate bias. They were given additional information on the

faculty and the number of graduates so as to give them each a similar base of information, beyond what they already may know based on the program's reputation (Jones, 1982). However, the names of the faculty can also increase the "halo" effect, that is, the impact of a single prestigious individual can inflate the rating for the entire program (Jones, 1982). The greatest service that the authors did to this study was to expand it to rely on much more than the reputational survey. The five other aspects in which a program was examined dilute the overall impact of the reputational standings, while at the same time enhancing the credibility of a reputational measure when it correlates to other sources of data.

The next general topic, "University Library Size," has but a single dimension. This data is from the Association of Research Libraries (ARL) and is referred to as the ARL Index. It is a composite score based on a variety of factors that are components of a library such as expenditures, number of volumes, periodicals and subscriptions, and staffing. The data only existed for 89 of the 228 schools that were involved in this study and in several cases the accuracy of the data is in question. The investigators were forced to rely on the ARL for this information because the expense of an additional survey was quite high; in return the data they received, while useful, did not include enough schools to have a great deal of influence on the overall ratings.

Two dimensions are examined under the topic of "Research Support": grants and expenditures. The more money that is put into research, the more research will be done.

More research does not necessarily mean better research, however there can be a wider breadth of investigation into a greater number of topics is available. Besides giving the university a greater direct impact in the scientific community, there are more opportunities for the student. Federal grants are a primary source of funding for research in the biological sciences and the focus of the first measure in this section. The rationale behind this measure was that the greater the quantity of valuable research, the more grants would be awarded. The second measure in this section examines the total expenditures of the university for research. Both of these measures rely on the assumption that more money will solve shortcomings in a research program. In reality, creativity, hard work, and innovation are all-important components of a successful line of research. However, the money helps, especially if it is efficiently managed and thoughtfully spent.

Courtesy of information from the National Institutes of Health, the National Science Foundation, and the Alcohol, Drug Abuse and Mental Health Administration, faculty members who held or previously held grants from these organizations over the course of the last three years were identified. The funds received from private sector grants as well as other public institutions such as the Department of Agriculture were not included. However, the investigators estimate that of all the faculty given federal grants, only 19% received funding from a federal agency other than the National Institutes of Health or the National Science Foundation. There are other factors that go into receiving federal grants besides the content of the research. These factors center around the faculty member and cannot be easily measured. Previous work, prestige, even grant writing skill

all enter into the equation, none of which does this measure address directly (Schwartz, 1992). In some ways, since grant awards are decided by peers, this criterion is another reputational measure. However, those individuals who decide on grant dispersal have a great deal more information on which to base a decision than just the reputation of the person or persons involved unlike the faculty members who are sent surveys.

The monetary expenditures of an institution on research can also be a useful tool in assessing the university's commitment to a program. The National Science Foundation was provided with this information on expenditures by the school itself in fiscal year 1978, including data from federal and non-federal sources (Jones, 1982). However, this data is not very useful in that the NSF only collected information on overall amounts spent for various areas, not individual programs. For example the data shows the total expenditure for biological sciences, but not a breakdown into each particular discipline. The other prominent difficulty with the use of this second-hand data is there is no control over how the universities reported their usage of funds. Each school could have a different procedure for internal accounting, thus confounding this data even more. In essence, this measure provides a good idea of the overall amount of research activity for a school, but is useless as a tool in differentiating between graduate programs.

The final area that the 1982 study examines is the publication record of the university. Two dimensions fall under this category. The first counts the number of publications attributable to the program in 1978-1979. The second attempts to measure

the "overall influence" of these articles based on the journal in which they are published. These measures work on the presumption that the more a program publishes, the more robust the research program must be. Both are based on data from the Science Citation Index from the years 1978 and 1979. The source files from the Institute for Scientific Information were processed by Computer Horizons, Inc., (now Computer Horizons Corp., NASDAQ: CHRZ) which designed the data processing program.

The first measure is the total number of articles that are attributed to the program in question. Association of an article to a discipline was based on the nature of the journal in which the article appeared. For example, articles that appear in the *Journal of Biological Chemistry* would be attributed to the biochemistry programs of the contributing universities. In journals that accept articles from all fields of science, credit is assigned based on the proportion of that field's contribution to that journal. In a case in which more than a single university has a contributing author, each university receives a proportional amount of credit. However, the data was collected by program and not by author. Therefore, a program that is responsible for an article with coauthors from within that program would get credit for a single article, whereas each program would receive credit for half of an article if the authors were from two different programs.

The second measure attempted to ascertain the influence of the published articles based on the prestige of the journals in which they were published. A count of citations in a journal was the determinant of the influence of a journal. If articles from a particular

journal were cited more frequently, then that journal was considered more influential in the field. Therefore, if a program published in a more influential, that is a more frequently cited, journal, that article was more heavily weighed in determination of the program's final score in this category. The authors make a point that this measure is not a sign of the individual authors, but an indication of the influence of the journal in which the work is published. The impact of an individual is not examined.

There are some difficulties with these measures as can be expected. The first is that neither of these measures include books published by the scholars of the programs. Since the primary means of information exchange in biological sciences is the journal, as opposed to the book, this point may not have much impact. However, a faculty member who devotes a large amount of time and energy into authoring a book or textbook goes unnoticed by this study (Jones, 1982). This measure also does not account for versatility in the faculty members. If an individual publishes in a journal outside his field, his program will not get credit for it. For example, if a faculty member whose research is focused on intracellular information exchange, is a part of the cellular/molecular biology program in a university, publishes an article in *Biochemistry* his own program will not get credit for it, but the biochemistry program at his university will. The measures also do not take into consideration the size of the program. The authors state that some thought was given to reporting this statistic in the form of articles per faculty member to help account for size, however this idea was rejected since the programs were given credit for all articles published, whether the author was a faculty member, postdoctoral apprentice,

graduate student, or any other individual affiliated with the program. The authors felt that people at various professional levels contribute to the quality of a program, therefore an overall program score would be more appropriate as a measure in this study (Jones, 1982). None of these criticisms cripple these two measures, however they do provide a perspective as to how much weight that they should be given in the overall evaluation of a program.

Examination of the available information yields a number of interesting points, all of which contribute to the idea of quality. Resulting averages vary greatly from discipline to discipline. Measure two, the number of graduates in the last five years, ranged from 14 in physiology to 26 in zoology. Measure eight, the quality of the program faculty from the reputational survey, ranged form an average of 2.7 in zoology to 3.0 in botany and physiology. Alone, these data are interesting statistics to have about a graduate program, but they only begin to reflect the idea of quality when compared to the other data collected. The authors examined the correlations between these points and the other data that were collected. Since all of the measures contribute to the quality of the graduate program, it follows that there should be a correlation between measures. The Pearson product-moment correlation coefficient was derived for each possible interaction between two points (Jones, 1982).

The first measure examined in detail for its correlation with other measures was Measure two, number of graduates. The authors do not simply assume that a greater

number of graduates means a higher quality program; instead they look at that statistic in light of the other available data. A high correlation between the various measures is a much better indicator that they all contribute to overall quality. As expected the number of graduates correlates highly with the other two measures of program size. Uniformly positive correlations were also found with the results of the reputational survey (Measures 8-11), research support (Measures 13-14), and publication records (Measures 15-16). The reputational survey (excluding Measure 10, improvement over the last five years) had exceptionally strong correlations in all disciplines expect botany, which, though still positive, did not achieve the robust .40 to .68 coefficients seen in the other fields (Jones, 1982). This high correlation shows that the programs that were the largest and had the highest number of graduates were more well known and received higher survey ratings. They also received the most support and had the greatest publication activity.

The second measure examined in detail was Measure 8, the quality of the faculty. It is no surprise that the reputational quality rating correlates highly with the size of the program. In addition, as expected from the Roose-Andersen and Cartter studies, there is a very high correlation with Measure 9, effectiveness of program, with none of the coefficients below .95. A similarly strong correlation is present with Measure 11, familiarity with a program, indication that respondents held the faculty of programs with which they were familiar in the high regard. There was a strong correlation (.47 or higher for all disciplines except botany with .31) between survey results and Measure 4, fellowship or training grant support. This could indicate that a program with a higher

quality faculty is capable of attracting more training money for its students. There was a strong (.23 to .36) correlation for Measure 6, commitments for employment, and an even stronger one (.30 to .63) for Measure 7, commitments for employment in an academic program (Jones, 1982). Perhaps this relationship indicates that students who have experiences with high quality faculty feel more comfortable in a similar role as a career. Positive correlations for both research support (Measures 13-14) and publication records (Measures 15-16) also make sense since faculty achieve reputation in some part through their publications and prominent investigators tend to attract more funding.

The third measure for which the authors examined the correlations was Measure 14, expenditures by the university for research in a discipline. As previously discussed, this data was from university expenditures for the whole of biological sciences; figures for specific programs were, for the most part, unavailable. Strong correlations (.42 to .78) are present for survey results (Measures 8, 9, 11), library size (Measure 12), and publication records (Measures 15-16) (Jones, 1982). Higher quality faculty will be attracted to a university that has ample research funding. It is also reasonable to conclude that the more funding a university has, the more it can devote to its library. Finally, if more money is devoted to research, then ideally more research gets accomplished, and hence, more articles are published.

The last detailed examination of correlations was made on Measure 16, "influence" of published articles. Naturally, there was a very high correlation (.90+) with

Measure 15, number of published articles attributed to a program. There were also moderate (.37 to .58) and strong (.56 to .85 respectively) correlations with Measure 1, number of faculty, and the reputational survey results (Measures 8, 9, and 11). This also makes sense, in that a program with more faculty or higher quality faculty is more likely to have articles published in prestigious journals. There was also a moderate to strong correlation (.37 to .78) with Measure 12, library size, and Measures 13 and 14, research support, with a single exception (Jones, 1982). This result points to the conclusion that the greater the support infrastructure, the more productive the program in generating quality research. The exception is botany's .04 correlation with Measure 13, fraction of the program faculty holding research grants from federal agencies. The only agencies included in this survey were the National Institutes of Health, the National Science Foundation, and the Alcohol, Drug Abuse, and Mental Health Administration, none of whom devote tremendous funds to botany (Jones, 1982). Had the Department of Agriculture grants been included this measure the results would have been very different.

It is important to realize that any single correlation is no more indicative of quality than any of the individual data points. However, when the results are approached as a whole, a general indication of how to better define a quality program develops. Naturally, it is a far more complex interaction than:

(Number of Faculty + Amount of expenditures) * (Library size + Reputational Prestige)= "Influential" Articles

Many factors that could play a part, from writing talent of graduate students to computer

and equipment resources to faculty undergraduate teaching loads, were not even measured. Beyond that many of the measures were incomplete, such as library size, or not completely equitable, such as not adjusting for size of a program in measures of publication records. However, study is useful to identify general trends that contribute to quality. If a small program wishes to improve itself, yet does not know where to devote its funds, it can use this study to demonstrate that a greater number of faculty can contribute to productivity, reputation, and quality, or additional research funding can enhance number of publications and prestige, and budgets can be built around these possible goals.

Despite the authors' endeavor to have this study accepted as a whole, it is natural to compare the reputational survey portion with studies that have come before. To forestall some of the previous criticisms that the other studies had faced, several adjustments were made to the protocol for this study. The most profound was the inclusion of data about the number of graduates and the names of the faculty members of the program. This additional information was added to 90% of the survey forms sent out, the remaining 10% were sent without extra data so as to have a basis for comparison with previous studies. Unexpectedly, the average rating for programs was lower when the respondents were given the faculty lists for the majority of the disciplines examined. The authors postulated that a respondent, when confronted with a large list of faculty whose names went unrecognized, would tend to not rate the program as highly (Jones, 1982). The differences were unlikely to have a major effect on the overall ratings because data

from both sample groups, with and without the extra information, correlated rather highly.

The next major addition to the survey was the addition of a question examining the familiarity of the faculty member with the program in question. This new question allowed the authors to investigate the effect of familiarity on the program's ratings. Approximately one in every seven responses indicated "considerable" familiarity with a program, most respondents said they had "some" or "little or no" familiarity with a program. However, when a comparison was done, there was a high (.77+) correlation between average ratings, though there were very few "considerable" responses on which to create averages (Jones, 1982). Much like the inclusion of the additional information on faculty, the assessment of data based on level of familiarity expressed in the survey question is responsible for some small changes in the ratings, but have little impact overall.

Critics have long argued that in the eyes of the public, visibility equals quality. The respondents were given the "Don't know well enough to evaluate" option with all of the questions and this option was often selected. The purpose of this option was to allow the respondents to avoid unfairly rating a program based only on the vague information that the subject might have available. Highly-rated programs tended to be highly visible ones. For example, only about 15% of the respondents exercised the "Don't know" option for schools with average ratings of 4.0 and up on Measure 8, rating of the program

faculty (Jones, 1982). Research work of high quality achieves high visibility for both the investigator and the investigator's institution. The important point to remember is that high visibility can often mean high quality; however, lack of visibility does NOT necessarily mean low quality.

Analysis was also done on the rater's alma mater and the geographical proximity to the other institutions in the area. After all, if a large program has more graduates, it is likely that a higher number of those graduates will be participants in the survey than a smaller program. Though alumni did give their former institutions higher ratings, the overall impact was small. Of all programs, 93% (574 programs) had their average rating unchanged and but a single physiology program had a change as high as .2 based on alumni bias. As to the question of geographical proximity, ratings tended to be slightly higher for the majority of disciplines when the rater was in the same area as the program in question (Jones, 1982). Because the investigators already demonstrated that respondents tended to rate institutions that they are familiar with more highly, it follows that smaller, less well known schools probably received the majority of their ratings from geographically nearby raters.

Though this study was the most wide-ranging ever done, in terms of different contributors of quality, there are still improvements that could be made. One such enhancement, of which the authors were strong proponents, was the inclusion of individuals outside of academia in the reputational survey. Industry is becoming a more

prominent employer of individuals with Ph.Ds in today's society. In order to be of value to the company for which they work, these individuals have to remain up to date with current research and advances and familiar with the institutions where the research is done. As active participants in the research community, the inclusion of these scientists from both corporate and government industry would be valuable in extending the scope of the reputational portion of the study. This group would also provide a different perspective on graduate education than the overwhelmingly academic point of view that dominates most of these surveys. Gone would be the concerns about current institution of employment. The question of how well graduate school prepared a student for life outside of the educational environment could be examined. However, this enhancement was not possible for this study because of funding restrictions.

Another area that could be improved involved the publication counts, specifically the "influence" measurement. As opposed to using a count of citations of a particular journal to determine the influence of that journal, a more direct citation count of the articles in question that are attributable to the school might create a more accurate picture of the impact of that article. More precision would be possible in properly attributing articles that faculty published in a field other than the program with which they are affiliated. There are difficulties with this approach, of course, sometimes the influence of a work is not evident until quite some time after its original publication or an article could have a profound effect in a small, highly specialized area of research. Issues such as these would have to be worked out before a system different from the one in this study

was used. However, due to both time and financial constraints, further exploration of this topic was curtailed for this study.

Different data collection techniques could have benefitted several measures, especially the research funding dispersal and library index. In the case of the monetary support, only aggregate numbers for the entire biological sciences funding were given. A breakdown by individual program would have rendered this data far more valuable to this study. Though all schools have different accounting mechanisms, some method could have been researched to provide a more accurate picture of how funds were dispersed among each department. This process, however would have been very expensive for the investigators and would also have required the cooperation of financial officers at many levels within each institution. Institutional cooperation would have been necessary had a more direct collection of data on libraries been attempted. As opposed to using the Association of Research Libraries Index, which represented only 89 of the institutions included within the study, a direct solicitation of information from the library in question could have been done (Jones, 1982). The response, hopefully, would have been greater than 89 out of 228. However, to pursue this course of action, additional resources would have been necessary to design an appropriate data collection instrument, send it to the schools, and process the data once returned. Both of these areas of investigation would have required the allocation of considerable time and money that was not available. Instead the investigators choose to use alternate sources of information--the financial reports and the ARL Index--to supply the data for these two measurements.

A minor improvement could be made to Measure 13, Federal Research Grant support, by including other Federal agencies besides the National Institutes of Health, the National Science Foundation, and the Alcohol, Drug Abuse and Mental Health Administration (Jones, 1982). The Food and Drug Administration, the Department of Defense, the National Aeronautics and Space Administration, and the Department of Agriculture are four such agencies that provide grants in the biological sciences that were not included in the original study. The topic of botany was especially under represented in its federal funding data because grants from the Department of Agriculture were not included. In addition, grants from private or corporate funds could be included in this measurement, though to acquire that type of data, the investigators would have to rely on the cooperation of financial officers within the school. The additional data would help to paint a better picture of how well a graduate program is funded.

In previous studies, the criticisms centered around the reliance on the reputational measurements as the primary source of data for the investigation. In this study, the authors sidestepped that criticism by having a multidimensional approach to examining the graduate programs. As opposed to being condemnations of the methodology, the comments centered around ways to enhance and widen the data collection process. This change could indicate that multidisciplinary approach taken by Lyle Jones, Gardner Lindzey, and Porter Coggeshall is a valid methodology to examine graduate programs in the eyes of the scientific community.

Contemporary Times (1993-Present)

Just as others before, the authors hoped that their study would be but one in a recurring series, in which each new study would build upon the old ones and enhance them based on both criticisms of others and novel ideas that a new investigator might have. Further, a new study could be compared to its predecessors and changes in graduate education could be charted over time. Finally, a study with current data would enable policymakers at schools to make better-informed decisions and allow prospective graduate students to make the best selection possible among the many programs available. Due to these reasons, in 1993 the Conference Board of Associated Research Councils commissioned the National Research Council to update the 1982 study.

Marvin Goldberger and Brendan Maher co-chaired the committee to conduct the study, which was published in 1995 under the title *Research Doctorate programs in the United States: Continuity and Change.* There were several primary goals for this investigation, first among them was to conduct another reputational survey to acquire the current views of scholars on graduate programs and compare these to the 1982 study. The second goal was to replicate and enhance the objective measures that were examined in 1982 (Goldberger, 1995). These two goals were the foundation for doing another multidisciplinary study that examined graduate programs on a variety of levels. The remaining three goals, comparison of the data, creation of a database that would allow others to use the collected data for personal research projects, and dissemination of the

report in an easy to understand manner, round out the investigators hopes for this project and its results.

The committee began, just as others had, by deciding on which fields of research would be included in this study.[7] These decisions were based on the number of doctorates awarded in each field annually, the number of programs within a particular field, and the number of doctorates produced by each program. Outside of biological sciences, all of the fields that were examined in the 1982 study were included as well as nine new ones. The additional fields help to address the changes that hadbeen occurring in graduate study and the wider diversity that exists in programming since 1982, as well as recognizing that more doctorates are being produced now than in the early 1980's.

The biological sciences presented a special problem to the designers. Over the decade between the two studies, the overall field of biological sciences changed radically with the emergence of new fields of study driven by techniques unavailable in the 1970's and the creation of interdisciplinary programs that drew upon more than one field to create a new speciality. The 1982 study included Biochemistry, Microbiology, Cellular/Molecular Biology, Botany, Physiology, and Zoology. Using the Doctorate Records Files of the National Research Council, the designers determined that there were fields that awarded more doctorates than Zoology, the topic with the least number of

---

[7]The methodology for Goldberger/Maher's is drawn from their 1995 work, *Research Doctorate programs in the United States: Continuity and Change*.

Ph.Ds awarded between 1986 and 1990. Almost three thousand degrees fell into these categories and it was clear that some restructuring was necessary.

The new fields were Ecology, Human and Animal Pharmacology, and Neuroscience. All of these fields gained prominence in the eighties. Consequently, the number of doctorates earned in these fields grew to the point that they could no longer be classified as "other biological sciences" and it was necessary to accommodate them in the new study. Taking these new fields into account, along with the explosion of new genetics techniques led not only to an addition of new fields but an entire restructuring of the taxonomy of biological sciences. Including the new categories and reassigning the old ones allowed the designers to create the following list:

- Biochemistry and Molecular Biology
    - Biochemistry
    - Cellular/**Molecular** Biology[8]
- Cell and Developmental Biology
    - Microbiology
    - **Cellular**/Molecular Biology
- Molecular and General Genetics
    - Human and Animal Genetics
- Neuroscience
- Pharmacology
    - Human and Animal Pharmacology
    - Toxicology
- Ecology, Evolution, and Behavior
    - Zoology
    - Botany

---

[8]The emphasis of the program toward Cellular or Molecular studies determined its placement.

- Ecology
- Physiology

This new classification system caused some difficulty at the institutional level in that it was not a required part of the original mailing. The original package showed the new taxonomy as "optional" and instead listed 11 different fields. These fields were condensed down into the seven fields that are listed above with the additional subdivisions to aid the institutional coordinator in assigning a program to a category and sent out in a supplemental mailing about a month later. Coordinators who had problems assigning programs at their institutions to the new classification had the opportunity to have their situation reviewed on a case-by-case basis. In all, 3634 programs at 274 universities were rated in this study, up by 965 programs and 46 universities. Biological sciences accounted for 974 programs at 206 universities were and 102 of these programs were neuroscience programs.

In pursuit of the stated goals, the designers created another reputational survey that they called the "National Survey of Graduate Faculty." A list of faculty members in each program was requested from the Institutional Coordinator of each school. If a faculty member held appointments in multiple departments or had a substantial contribution to the graduate education of more than one program the Institutional Coordinator was asked to list that faculty member in each program in which he or she was a part. From this pool of submitted scholars, the committee selected the recipients of the survey instrument.

In order for a university to be included, it had to fulfill certain eligibility requirements. A university had to produce at least three doctorates in between 1988 and 1990 and at least one doctorate in 1991, within a field. Alternately, if the program had received a rating of 2.0 or greater in the 1982 study, only the three doctorates between 1988 and 1990 requirement needed to be fulfilled. If at least one program in a university met these standards, the university was included both as an object of rating and a source of raters.

The survey instrument itself was a questionnaire that was very similar to the one used in the 1982 study. Personal information, such as alma mater and specialization within the field was requested, as was a recommendation of at least two individuals from foreign institutions that could act as raters for American graduate programs. The actual survey replicated the 1982 survey with one alteration. The question about familiarity with the program was split into two questions, asking the respondent to rate familiarity with the work of the faculty and familiarity with the graduates of the program separately. Beyond that change, the questionnaires themselves were identical. Raters were still provided with data on the number of doctorates produced and the names of the faculty members and the range of responses was the same.

There were differences in the presentation of the surveys, however. The goal of the committee was to have at least 100 responses for each program. The number of raters was set at four times the number of programs for each individual field, or 200 whichever

was higher. To reflect the large interdisciplinary environment of the biological sciences the number of raters was increased to five times the number of programs within a field. Therefore each program was included on at least 200 surveys, unless it was a biological science program in which case it appeared on at least 300 surveys. A total of 16,738 surveys were sent out to scholars in all fields, of which 5,331 were to members of the biological sciences community. Approximately 7900 raters responded before the February deadline, overall with 1967 usable responses in the biological sciences. Each biological science program met the goal of at least 100 ratings and the overall response rate for biological sciences was 43%.

The investigators were also very careful to include a broad spectrum of raters. The country was divided into nine geographical regions including between three and nine states and territories. Great care was taken to insure that surveys were sent to faculty members in all nine regions. Useable responses ranged from 28% in the New England region[9] to 65% in the West South Central region[10]. The regions that were sent the fewest had the highest percentage return of useable responses. Care was also taken to vary the raters by academic rank so as to include raters other than full professors. Junior faculty accounted for 2121 recipients of the survey and they had a slightly higher percentage return of useable responses then did full professors.

---

[9]Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont

[10]Arkansas, Louisiana, Oklahoma, Texas

Neuroscience had a very good percentage return as compared to most of the other fields. Full professors, associate professors and assistant professors had useable return percentages of 45%, 49%, and 57% respectively. This was the highest percent return of all biological sciences fields by all ranks, with one exception. Biochemistry had a slightly higher 58% useable return rate among assistant professors.

There were several other manipulations done on a small group of the surveys. Ten percent of the questionnaires replicated the directions as they were given on the 1982 reputational survey. This group also did not include the new question regarding familiarity with program graduates. There was, however, no significant in the pattern of responses, between the recipients of the "old" instruction set and the "new" instruction set. Another question asking the respondent to indicate the number of faculty in a program whose work was familiar was also added to ten percent of the questionnaires. The level of familiarity can be used as an index of visibility of the programs in question in later comparisons

The investigators wanted to examine the idea of expanding the survey to include individuals in non-academic settings just like the 1982 group. The committee did a brief study to assess whether a "National Survey of Industrial Employers" would be possible. Information was gathered at several organizations that focused on electronics and the investigators concluded that it would be possible to accomplish a survey in this field of industry. Additional resources would have to be allocated to design and implement this

survey, however, they were unavailable and, therefore, the industrial survey in electronics was not done. Nor was any further exploration into the possibility of an industrial survey in other fields. However, just as before, the authors stressed that this area must be examined in the future because of the ever greater percentage of doctorates employed in an industrial environment.

The objective measures underwent considerably more changes than the reputational study. The five different categories of the 1982 study were condensed into three, old measurements were modified and reshuffled, and several new measurements were created. The categories of "Faculty," "Students," and "Doctoral Recipients" were defined and measures assigned to them for each program. Beyond the measurements taken for the programs a series of categories was designed for the university itself, to address descriptive statistics that could not be attributed to a program alone, but rather applied to the entire institution. "Institutional Profile" is a category that gives general information about the university, "Research and Development Expenditures" examines the amount of financial expenditures of the university as a whole, "Library" gives information on the institution's library resources, and "Enrollment" gives data on overall number of students. Between these two areas of measurement, the program level and the institutional level, almost all of the objective points from the 1982 study are used, though some were modified. The two points that were omitted were Measures 6 and 7, which examined the commitments to employment that students had made prior to graduation.

The first set of measures focuses on the aspects of graduate education at the program level. The first category being Faculty, has eight dimensions[11] that incorporate measures from three different categories of the 1982 study. Aspects of "Program Size," "Research Support," and "Publication Records" are all included under this umbrella. The first dimension is Total Faculty, provided by the Institutional Coordinators. In addition to the total number of faculty the Institutional Coordinators were asked to provide the percentage that held the rank of full professor, that percentage became the second dimension. The third dimension, mirrors Measure 13 from the 1982 study, asking for the percentage of faculty who had been awarded federal research grants between the years of 1986 and 1992. Addressing the previous criticisms, grants from other federal agencies were examined, adding the Department of Defense, the National Endowment for the Humanities, the Department of Energy, the Department of Agriculture, and the National Aeronautics and Space Administration to the list of federal supporters with the National Institutes of Health and the National Science Foundation.

The remaining five dimensions examined all relate to publications. The data on publications and citations from the Institute for Scientific Information was much more accessible than in the previous study so a new group of measures was created to replace the old Number of Publications measure and the problematic "Influence" measurement.

---

[11]Only five dimensions were present in Arts & Humanities. Instead of data on publications and citations, awards earned by both the program and the faculty were included due to the different nature of disseminating information in the non-science fields.

The first new measure was a straight percentage of the faculty publishing between 1988 and 1992. The second new measurement was merely a derivation of the first, dividing it by the total number of faculty to arrive at a ratio of publications per faculty member in a program. To combat an artificially high ratio the investigators also included a Gini coefficient[12] for publication as the third measure. This number is an indicator of how the number of publications is distributed among the faculty of a program, a high number indicating that a few of the faculty are prolific publishers. The fourth new measure is a ratio of citations to the total number of faculty. This measurement is a follow-up to one of the suggested improvements from the 1982 study. With the advent of far more efficient data processing technology, it became feasible to match up more than a million citations to members of program faculty included in this study, replacing the "influence" measurement with a far more precise measure. The final measurement was a Gini Coefficient for program citations. Computed in the same manner as for publications, this number indicates whether only a few researchers are cited from a program or if all of the faculty have about equal influence. However, it is important to keep in mind that highly specialized fields of research may only garner a few citations, but may be groundbreaking nevertheless.

The Students category contains only three dimensions derived from the "Program

---

[12] This statistical manipulation indicates the concentration of publications to a few individuals within the program. If only a single individual published in the program the Gini Coefficient would be 100, if everyone in the program had an equal contribution the Gini Coefficient would be 100 divided by the number of program faculty.

Size" category in the 1982 study. The first dimension, Total Students, indicates the number of graduate students enrolled in a program in the Fall of 1992. The second dimension is a subset of the first, indicating the percentage of students enrolled who are female. The final dimension is the reported number of Ph.Ds produced by a program between the 1987-1988 academic year and the 1991-1992 academic year. The Institutional Coordinator provided the data for all three of these measurements.

The last category on the program level, Doctoral Recipients, had six dimensions drawn from the "Characteristics of Graduates" section of the 1982 study as well as adding some descriptive statistics about the program as measures. The data for all of these measures was acquired from the Doctorate Records Files (DRF) provided by the National Academy of Sciences/National Research Council. The first three measures are simply descriptive statistics listing the percentage of the Ph.Ds that were awarded between July 1986 and June 1992 to females, minorities and United States citizens or permanent residents. The next two measurements examined the funding of graduate students, questioning the percentage of Ph.Ds who were supported primarily by research assistantships and the percentage supported by teaching assistantships. The final measure is the same as Measure 5 from the previous study, median years from first enrollment in graduate school until receipt of the doctorate.

The institutional level objective measures examine the universities as a whole, including aspects that cannot apply to a single program alone, but rather to the

environment in which the program exists. The data collected were divided into three categories along that each described a portion of the university. "Institutional Profile" that gave information about the type of university, "Research and Development Expenditures" that gave information on funding, and "Library" presenting data on the library resources were the categories, coupled with information on the enrollment of the university and the number of programs in both the 1982 and the 1993 study, made up the presentation of data on each school in the study.

The "Institutional Profile" gives descriptive information about the university in question. The first area examined is the year that the first Ph.D. was awarded. The DRF is the source of this information from 1920 on, prior to that however, other sources were utilized. The next area that is examined is the Carnegie Classification[13] of the university. Carnegie Classifications are labels for institutions assigned by the Carnegie Foundation for the Advancement of Teaching that provide a description of the institution's mission and are based on the number and type of degrees offered among other factors. The final factor in the profile is the type of "Institutional Control," that is, is the institution a public or private school. This information was provided by the US Department of Education.

The area of "Research and Development Expenditures" presents information on the money spent at the university. There are two dimensions to this area, Total R&D and Federal R&D. The former is the average annual expenditure from 1986 through 1992,

---

[13]For a description of the various Carnegie Classifications see Appendix I.

the latter is the average federal expenditure over the same time period. The data was provided by the National Science Foundation and was corrected for inflation and presented in 1988 dollars. Much as in the 1982 study, the same criticism can be levied at this data, that it would have been far more useful to see expenditures in the program-specific portion of the objective measures. However, this was one improvement that was not acted upon and will have to wait for the next study.

In 1982, a major criticism was of the use of the ARL Index as a source of data on university libraries. Not that the ARL Index was a bad source of information, it just did not have data on the majority of the schools in the study. The 1993 investigators took this criticism to heart and expanded the sources of information to include not only the Association of Research Libraries, but the Association of College and Research Libraries and the Department of Education as well. Three different elements of data were collected on each library. The first was the number of volumes present in the 1992-1993 academic year. The second was the number of serial subscriptions held by the library for the same time period. The third was the total funding expenditures for 1992-1993 from all sources.

This data, along with figures on the total enrollment of the university and the number of graduate students enrolled, provides an overall picture of the institution in which a program has to function.

The findings from this study are consistent with those of the 1982 study,

determining a number of strong correlations between the objective measures and the reputational study. The "Scholarly Quality of Program Faculty" portion of the reputational study is consistently the most highly examined portion of this type of study and the 1993 study is no different. The authors are careful to point out that just because a program rates highly on this portion, they might not have such positive results in other portions (Goldberger, 1995). However, nearly all of the data is presented based on the rank order established by this question. On this question, about 43% of the programs rated as strong or distinguished (3.0+) and about 19% were rated as marginal or not sufficient (less than 1.99) for graduate education across all of the programs. In the biological sciences, the numbers were very similar at 45% and 19% respectively. On the question examining the "Effectiveness of Program in Educating Research Scholars/Scientists," 21% of all programs were considered extremely effective (3.5-5.0) and 9% were rated as not effective (less than 1.49). The biological sciences had slightly better than average percentages for both categories at 24% and 8% respectively (Goldberger, 1995).

The objective measures support, with positive correlations, the general trend of programs in the reputational study. A number of the "Faculty"measurements can be compared to the question of "Scholarly Quality of Faculty," in that they should be strongly associated. The first is the federal grant support in the sciences (Goldberger, 1995). A higher quality faculty member should be awarded more in federal grants, and a program with higher quality faculty should have received greater federal support. In the

top quarter of programs in the biological sciences, as determined by the "Scholarly Quality of Faculty" measure of the reputational study, the majority of the faculty had some federal support between 1986 and 1992. Ecology, Evolution, and Behavior had the lowest percentage of federally supported faculty at about 58%, whereas both Neuroscience and Biochemistry & Molecular Biology reached a high of about 77% (Goldberger, 1995).

Another set of measurements that should be strongly associated with faculty quality is the publication and citation category. A higher caliber of faculty would publish more prolifically, and their work would be cited more often. In the biological sciences, Neuroscience was selected as the topic on which a more detailed analysis of publication records was done. In all four quarters at least 81% of the faculty had published at least once between 1988 and 1992, ranging from 81% for the fourth quartile to 89.62% for the second quartile (Goldberger, 1995). A trend appears in the next measurement, publications per faculty member in the program. The top quarter had a score of 11.97, over twice as many publications per faculty member as the fourth quarter with a score of 5.21. The same trend was detectable in the citations per faculty member, with the top quarter scoring 136.51 cites/faculty member, almost twice the score of 69.4, held by the second quarter and almost seven times the score of 20.26, held by the fourth quarter (Goldberger, 1995).

Size is yet another correlate of program quality, both in size of the faculty and

student populations. Higher rated programs tend to have more faculty, especially in the biological sciences in which many of the programs are interdisciplinary. The top quarter of biological sciences had an average faculty of 50, the bottom quarter had an average of 10. The number of students at the top quarter of schools averaged at about 59, and at the bottom about 14 (Goldberger, 1995). Higher quality programs tend to attract more students.

The investigators present all of the data collected in a manner that they hope will be easily accessible to other researchers. The analysis conducted in this study only scratches the surface of the data, and its potential uses. Marvin Goldberger and Brendan Maher, just as their predecessors, suggest that this study should be one in a series, done periodically to measure the current state of quality in graduate education. They are strong proponents of expanding the reputational study beyond the academic environment into the realm of industry. They also believe that the objective portion should be expanded to include the careers of program graduates. Funding was not available for either of these enhancements, but they should be important considerations for the next set of investigators to pursue this legacy of studies.

Changes for Tomorrow

All of these studies provide insight into how quality can be defined. Hughes was the first to say that other scholars, learned in the subject could provide an index of quality. This initial assertion has been modified and added to through the years with many enhancements. Keniston expanded the investigation to examine more institutions in varying fields. Cartter added comparisons of descriptive statistics such as library resources and salary to the faculty survey. He improved the survey portion by enlarging it to more than four thousand individuals and conducting numerous internal analyses for bias on topics like alma mater and geographical location. Roose and Andersen expanded the format of Cartter's study to include programs in more universities and in more fields. The 1982 study led by Jones, Lindzey, and Coggeshall revamped the use of objective measures and attempted to reduce the influence of the subjective faculty survey by making it a single subset of measures among five others. Finally, in the 1993 study, Goldberger and Maher updated the format of the 1982 study, reshuffling the objective measures and adding a few new ones. From these studies areas of importance can be derived and within these areas improvements to a program can be made.

Improvements vary from altering small details to profound changes in the structure of the program and cannot all be addressed on a single level. A multi-tiered approach that incorporates changes at the level of the university, faculty, the student, and the graduate program is necessary, altering at the level to best achieve results.

Enhancements to a program can directly impact of survey areas or they can be subtle and affect a survey question only indirectly. Yet, the studies of graduate school quality do give an idea of major topics that are components of quality and improvements should strive to better these components, no matter what the direct impact or the level on which it occurs. The starting point that is given by the indices of the studies provide a concrete basis around which a program director can plan improvements.

The university both directly and indirectly contributes to graduate education, but these contributions are as essential as any other. A university, that is, the institution within which a neuroscience graduate program exists, is the component of graduate education that seems the most remote and unaffected by the actions of a student. A university is generally an enormous entity that provides space and resources to a graduate program, yet in a manner that is beyond the knowledge of the average student. Usually interactions with the university are done by the program and the student need not be involved. These interactions can result in resource allocation, funding, personnel, and use of shared assets. Study indices regarding the availability of programs, library resources, Carnegie classification, and total research and development expenditures are a reflection of the university as a whole, often beyond the scope of an individual program to control.

United under the banner of a university, a faculty shares the common responsibility of educating students in not just the hard science and factual information and the logistical aspects of research, but in how to think. The execution of an

experiment and the collection of data can take a great deal of skill, but interpreting the data, conceiving and designing the experiment, and knowing its place in the larger scheme requires the ability to think independently and creatively. This talent is the key separating factor between a scientist and scholar and a highly skilled technician. The responsibility of teaching the student how to think effectively is a profound one. Each student is different, and the faculty must approach the student in a manner that will facilitate learning for that individual student. These responsibilities are examined in the indices that spring from the reputational portion of the studies. In addition, objective measures such as number of publications and citations attributed to faculty members and amount of federal research support are available. A process of improving faculty performance that is built around these points will enhance the quality of the program in the next study.

The university and the faculty both have responsibilities in improving neuroscience graduate study, but there is another important component in the equation, the graduate student. The student is the individual that graduate education is focused upon, the future scholar, scientist, and educator. That student will be molded and changed by the program in which he studies and the faculty with whom he studies. Should he work in the academic world, the former student will become the teacher and will impact the lives of the next generation of students. A graduate program is built around educating students in neuroscience, a university is built around educating students in a vast array of fields, and faculty are the instruments of this instruction. However, the

student is not simply a passive observer, the job of a graduate student requires active participation in the educational experience. However, the student is the least addressed by the points examined in the studies. Aside from financial support and the number of years that it takes to achieve the Ph.D the students are not a major determinant of the quality of a graduate program. The investigators hoped to expand the examination of the student by surveying the post-graduate work of the graduates of each program, however, they lacked the resources to do so.

The program is where all of the components of graduate neuroscience come together. The program is the part of the university charged with teaching neuroscience. It provides the environment in which faculty teach classes and conduct research and gives the student the resources to earn the Ph.D. A neuroscience program is the medium in which all aspects of education interact under the banner of neuroscience graduate studies. As an entity, the program has considerable responsibilities in education and administration, as well as responsibilities to the welfare of its students. The university relies upon it to be a vehicle to promote research and training in the neurosciences, the faculty relies upon it to provide an environment in which to work and students to train, and students rely on it for just about everything. Size of both the faculty and student body are examined in the objective portions of the studies, as is the number of PhD.s produced. The reputational portion also examines the program's effectiveness in training students and the improvement from previous studies.

Though the indices, both reputational and objective, do not give the complete picture of how to provide quality in graduate education they provide a beginning. A university that is dedicated to the improvement of its graduate program in neuroscience can use the results of these studies to plan changes that will have a positive effect on the quality of the university in the next version of graduate school ratings. These changes must be accomplished by the efforts of the university, the faculty, the student, and the program all working around the higher goal of an enhanced program.

The first component to this plan of improvement must be the university. The university must be willing to make a neuroscience curriculum a part of its academic endeavors. Before any of the specific indices from the studies can be applied to a program to define its quality, the university must take steps to define the program's existence. A commitment must be made by the school to devote resources, personnel, and money to creating and maintaining a neuroscience program. Unlike many academic disciplines that have existed an university departments for years, neuroscience is something new, most of its programs having been created in the last 25 year (Association of Neuroscience Departments and Programs, 1994). Naturally, when a university dedicated itself to a new program, that university will want the best program possible. However, some of the more basic issues of support and infrastructure must first be addressed.

The physical resources provided by the university include the space and utilities

that house the program. Lab space, faculty offices, administrative offices, classrooms, student lounges, and storage facilities are all assigned by the university. Utilities such as water, power, sewage, telephone and computer lines, and gas are also necessities for a graduate program in the sciences. Beyond these obvious physical resources there are other support services that the university provides. Examples include janitorial service, security, and animal services. It would be financially irresponsible as well as logistically difficult if each department hired its own cleaning staff and police. Universities will generally try to minimize the number of animal storage facilities on a campus to conserve resources as well as for security reasons. However, the space, utilities, and animal care staff are generally provided by the university for a per diem care charge to the program. This centralized animal facility is shared by several departments, and removes the burden of caring for the animal's day-to-day needs from the program personnel.

Personnel support is also provided by the university. The process by which faculty, technicians and staff are hired varies from one school to the next, but the university as a whole provides for support services for individuals who work within a program. These services include benefits like retirement investment and health insurance as well as payroll and tax withholding. It is more cost efficient for the university as a whole to administer the bureaucratic aspects of employee maintenance that are common for all the employees.

The program is only a part of the larger whole that is the university and just as

janitorial, security, and animal care are shared throughout the community, so is administrative support. The staggering amount of bureaucracy that is present in a university necessitates an efficient administrative body. Programs will have input on many tasks, but the "legwork" is done by another body in the university. Admissions is one example of this duality. The department will review the candidate and decide whether to offer the prospective student a spot in the next class. However, the gathering of information, organizing the student's file, and doing much of the paperwork falls to the admissions office. Other student services such as payroll, registration, and the bursar are handled in a similar manner. Another example of this phenomenon is the purchasing and receiving department. A department will decide what it wishes to acquire and from whom. However, the university then has guidelines for the purchase, receipt and delivery of goods. These procedures will vary from university to university, but they are necessary in order to run an efficient campus. The actual logistics of processing the purchase order, cutting the check, receiving the material and delivering it to the ordering entity falls to the purchasing and finance department of the university. Unifying these tasks under the umbrella of the university makes the school more efficient as a whole and saves each department the cost of performing these tasks alone.

The other type of resource not provided directly to the program or the student, but vital to both are common resources such as library facilities. The library is an essential part of the graduate experience. Attempts were made to collect statistics during both the 1982 and 1993 studies, using the ARL Index and individual questionnaires respectively.

The highly-rated schools had better equipped libraries (Jones, 1982). It is not just the size of the library however, but its contents that are important. That is why, in the 1993 study, the investigators expanded the study to examine the number of periodical subscriptions as well as the number of volumes (Goldberger, 1995). Highly rated schools had the more extensive collections, more books, and a greater number of subscriptions to periodicals. A school's scores in the various library indices could be enhanced by expanding the collections and increasing the number of periodical subscriptions. Unfortunately, according to recent statistics published by the Association of Research Libraries, expenditures on journals have been cut back by 8%, despite the fact that there are 24% more graduate students using these libraries than a decade ago (Shapiro, 1997). This trend must be reversed to increase the scores for the university as a whole, however, fiscal realities may dictate otherwise. In order to benefit neuroscience, the program would have to take an active part in providing guidance to the library at its university so as to maximize the use of available funds. Input and suggestions from the department would help the library to concentrate its funds on journals and resources that would be most advantageous to the neuroscience students and faculty.

The other indices of quality that were examined on the university level was the total expenditures on research and development and the federal expenditures on research and development over a five (academic) year period. Based on data from the National Science Foundation, this amount reflects the total amount spent across the whole university. The investigators had hoped for a more detailed breakdown, however that was

not feasible for this study given time and financial constraints. Highly rated universities tended to have higher expenditures, and the message to schools is clear (Goldberger, 1995). Spend more on research and acquire more federal funding, that is a hallmark of a quality university. Though this measure did not directly address expenditures on neuroscience, a neuroscience program can benefit from a general increase in research funding if some of the money is dedicated to neuroscience research. In addition, the interdisciplinary nature of neuroscience is a benefit, because many faculty hold multiple appointments in different departments and can reap additional funds that can contribute to neuroscience research.

The remaining indices that were examined in the 1993 study were simply descriptive statistics about the universities. These included the type of institution (public or private), year of the first PhD. award, number of students, and Carnegie classification (Goldberger, 1995). The first cannot be changed without legislation and the second cannot be changed at all. The number of students, both graduate and undergraduate fluctuates constantly and comparison of graduate student is better done on the program level. Changing the Carnegie classification would require a major drive for federal funds as well as alteration in the number of programs and doctoral awards. With a few exceptions, having graduate programs automatically places the school in one of four categories, Research University I or II or Doctoral University I or II (Carnegie Foundation,1997). The specific assignment is determined by the number of degrees awarded, the number of disciplines in which PhD.s are awarded, and the amount of

federal grants given to the university. To alter this classification would not as much affect a specific program, but the focus of the university as a whole.

The faculty are the second component to a plan of improvement to a graduate program. Unlike efforts by the university, that affect the entire school, enhancements that revolve around the faculty can specifically benefit a neuroscience program. The university provides the raw material for an endeavor into graduate level pursuits in neuroscience, but the actual sculpting of a program is done by the faculty. These people are the teachers, researchers, administrators, and mentors who give the program life. Whether leading a class or making policy, a faculty member will have a profound impact on the students with whom they come in contact. Faculty guide the students, interact with the university, and manage the program. The student will come into contact with the faculty in every aspect of graduate education, depending on them for guidance, knowledge, opportunities, and contacts in the scientific community. Faculty members can make or break a student's career, can change the direction of a program, and are the most influential component of quality as judged in most of the studies of graduate school quality.

The impact of the faculty is the central component of the reputational portion of the study. Questionnaires asking peers to rate a program by the scholarly quality of its faculty and the effectiveness of its program have been the heart of reputational studies as far back as Hughes' work. These reputational measurements have been expanded and

refined over the years, but they still represent the core of program quality rating. Determination of a program faculty's scholarly quality requires the rater to make a judgement based on what that rater knows of the program faculty members' research, publications, and merits. The determination of the program's effectiveness necessitates the rater's consideration of the faculty's ability to teach and produce skilled researchers, in essence, how the faculty influence the student.

The average beginning graduate student sees the faculty in a similar manner as when the student was completing undergraduate training. The faculty are the teachers, they lecture in class, they advise in coursework, and they assign grades. All of these are roles that they still fulfil, there are few neuroscience graduate students who do not have some basic classes as part of the curriculum. Anatomy, biochemistry, introduction to graduate neuroscience, cellular biology, physiology, psychology, and pharmacology are all classes that provide the student with the basic knowledge that will be necessary to pursue a graduate career in the neurosciences. Beyond science classes there are other foundation classes that will be necessary to the future researcher such as grant writing, statistics, and ethics. These classes are generally taught in the manner that is familiar to the student from the undergraduate environment, a teacher leading a class, notetaking, tests, papers, and a final. Graduate classes at this level can be shared between different programs, departments or even with students outside the graduate school. It would not be uncommon to find a neuroscience Ph.D. student, a biochemistry Ph.D. student, a cell biology Master's student, and a medical student in the same Biochemistry class.

Similarly, a graduate ethics class could have a broad spectrum of students from many graduate programs. These types of classes are taught with the faculty in the traditional role as teacher and class leader, not much of a change from teaching undergraduates.

As the student progresses beyond the basic classes, the faculty role as a teacher begins to change. An upper level class is smaller and more intimate, and the expectations are higher. The student will find that interaction with the faculty member is mandatory and being merely another face in the crowd is not an option. The faculty member, having a smaller class, is able to devote a greater amount of time to each student, and also to have higher expectations. These are the classes in which a faculty member can endeavor to bring out the talents of students and focus upon improving shortcomings. Classes in which the scholar forces the student to think, to be creative, and to interact and build upon the ideas of others are often the most successful and memorable of graduate school and a skilled educator can facilitate an upper level class so that it is not just a forum to exchange information, but to explore knowledge. A teacher who can accomplish this difficult task will have students lined up to take classes, no matter what the difficulty.

Once out of the classroom and into the laboratory, the faculty member becomes a teacher in a different manner. The student will not only listen to what is said, but watch what is done. The student will be watching every aspect of how a faculty member acts in the lab in the hopes of learning how to effectively lead a research group. Any lab in which the student works, be it as a rotation, a technician on a research assistantship, or as

a doctoral candidate, can potentially provide the student with knowledge above and beyond the technical skills associated with the line of research pursued in that lab. The faculty member should teach the student about the lab and the research conducted within, that knowledge is essential for the student in deciding what line of research will be pursued for the doctorate. However, there are many parts of working in a lab that are not covered in class or in the literature.

Logistical organization is necessary to run a successful lab. For as any faculty member knows, running a lab is not all research. Use of grant money, animal services, purchasing and payment procedures, proper timing of an experiment, and acquisition of office supplies, among many other details all play an important part in a successful lab. A faculty member who is an efficient manager is an asset to a lab and if those talents can be passed on to the graduate student they will be useful in that student's own career.

The faculty member will best teach the student in the lab by example. If the faculty member is enthusiastic, open, passionate, meticulous, and dedicated then those are qualities that can be passed on to a graduate student. However, the professor must make the time to interact with the student, to talk and process the collective efforts that the student observes and is a part of. To guide and nurture a student is not to coddle, however. A professor can make the lab experience challenging, so as to stretch the bounds of the student and force a new way of accomplishing a task upon him, but at the same time this challenge cannot exist in a void, the faculty member must be available to

provide support, criticism, and positive reinforcement.

The most profound form of this nurturing relationship is that of the mentor. This is the faculty member who accepts the responsibility of nurturing the graduate student through Ph.D. research. This faculty member will open up his lab, devote time, effort, resources, material and finances on the student on the quest for the doctorate, and will be expected to become more than just another teacher. The mentor-student relationship becomes a part of the life of both faculty member and student and should be entered into only after the most careful deliberation by both parties.

The faculty member will be called upon to devote many resources both professional and personal to the student. Within the laboratory environment, the student will be granted a new status. No longer a glorified technician, now the student is a co-investigator, attempting to develop a unique line of research about a topic similar to that of the mentor's. The mentor must make sure that there is room for the student to grow in the environment that he creates. There are logistical considerations to consider, is there enough funding on a grant to carry this student, is there enough room in the lab to provide workspace, is there any office space available? There are research aspects to consider, does the lab's current research have room for expansion, if there sufficient latitude for the student to develop an independent set of studies that could quality for a doctorate? Finally, there are personal aspects for the faculty member to consider, do I wish to devote the time and effort to this student, will we have a rapport, is the student worth a four to

five year commitment?

The responsibilities of a mentor are those of a teacher in the most extreme, for the mentor bears the burden of transforming a neophyte graduate student into an independent thinking researcher. The mentor will guide the student through selection of a topic of interest for dissertation research, planning the studies, performing the experiment, and writing it up in the proper manner. A skilled mentor will not give away answers, just ask questions that will prod the student on the path to finding his own answers. A balance must be maintained between allowing the student enough independence to learn and grow and providing enough of a presence for the student to rely upon in times of need.

In working with the mentor, the student will learn by example many aspects of commanding a research project. Laboratory management skills, staff and personal scheduling, resource management, planning and organization are all skills that a student can glean from a talented mentor. In return, the mentor should be prepared for questions of this nature, not scientific, but bureaucratic, for these skills are as vital in today's scientific workplace as any others. This does not mean that scientific questions are any less important. Selecting the proper topic alone should be the culmination of a great deal of preliminary research, discussions between mentor and student, and an analysis of resources available. The design of a study and each of its component experiments should likewise be the product of careful planning. The dissertation research period is the time when a graduate student can afford to make mistakes, the mentor is there as a safety

control, the trick for the student is to learn from these mistakes and how they affect the research process. Allowing a student to make an error can be difficult for a mentor, but it is sometimes more valuable to the student's development, though at a loss of time or resources.

A talented mentor is a rare and valuable commodity in a graduate program. Faculty members can be fine scientists, skilled researchers, and capable teachers but to be a mentor takes a far greater investment. The mentor must be able to see the big picture for that graduate student, and be willing to put the time and effort in to not only supervise the student but to get to know the student's strengths and weaknesses and develop a rapport that will promote the student's success. Students will recognize this type of individual and flock to him as a mentor if there are any aspects of his research that the student can match to his own interest. This is the type of person that a school should endeavor to attract and add to the ranks of the faculty.

A program which possesses a faculty that can excel in the challenges that accompany the training of graduate students will receive higher ratings in the reputational study. As doctorates of that program enter the world of neuroscience research, their conduct will reflect well on the program and enhance its reputation.

Beyond the reputational survey, statistics on publications were gathered in recent studies, most thoroughly in 1993. The number of publications themselves were

measured, as well as the number of citations of those publications, in reference to the number of faculty. The Gini coefficient was also used to determine if a program had contributions from a large number of faculty of if there were just a handful of prolific publishers. Highly rated programs have more publications per faculty member, as well as a higher percentage of the faculty who regularly publish (Goldberger, 1995).

In light of these indices, there are steps that a faculty member can take to increase the quality of the graduate program. To enhance the reputational ratings, the faculty member must strive to be a good scholar and researcher. Being well known to peers for talent in research will demonstrate the scholarly quality of the faculty. Publications also reflect the effort of the faculty in research pursuits and a high quantity of citations indicates that the researchers work impacts many other studies. The literature is an important medium to educate peers about the research, methodology, and ideas of the faculty as well as to promote name recognition. The faculty member must also be a good educator. The scientific community will judge this aspect of the program faculty by the new doctorates who enter the workforce. If these new researchers are highly skilled and talented, the program's reputation will be increased and higher ratings will result. If they are ill trained and less than competent, it will reflect badly on the program. These efforts by the members of the faculty are essential in improving the quality of a program, for the reputational survey is still the component of assessments of graduate programs which receive the most attention.

The third component to a plan of improvement is the student. The fact that a student must work hard in a graduate program comes as a surprise to no one. However, beyond the student's own experiences there are steps that the student can take to improve the program overall. Most students do not realize this possibility at first given how intimidating the initial graduate experience can be. No matter how friendly the faculty, no matter how flexible the program, the incoming student will not want to make waves that could potentially affect him over the next five or six years. But, there are many constructive ways that a student can contribute to enhancing the quality of a program.

Simply performing high quality work, a goal of most graduate students, will help the program. If a student publishes research, the visibility of the university is increased. Publication of quality work will create a positive reputation for the student. That favorable regard can be attributed to the school on questionnaires that ask scholars to judge the quality of graduate education at a university, such as in the studies cited. If a student who graduates from a program is a highly competent professional, an efficient lab manager, and an innovative teacher with several studies and publications on his resume, positive letters of recommendation from faculty, and favorable evaluations from students he has taught, he will have some degree of respect in the academic arena. This respect will be shared by the school that spawned this new Ph.D. Of course, the door swings both ways. Should a student be lackadaisical, lazy, or incompetent, the school's prestige will suffer. Either result can be seen in the reputational survey within the questions about the quality of the program in training students and the relative change in the program

quality.

Two of the objective indices examined a primary concern of the graduate student, money. With the cost of tuition having increased by five percent the last two years, (Gose, 1997) the financing of a graduate education can be an area of extreme concern to a graduate student. Not only must tuition and books be a consideration, but also living accommodations, utilities, bills, transportation, some occasional fun, and in some cases, a spouse and even children and financial aid is often a necessity. A common source of income is the student loan. Fifty nine percent of doctoral students reported borrowing an average of $19,245 (Basinger, 1998). Grants, loans, scholarships, independent wealth, and parents are other sources of income but the study only examined the percentage of students who had an assistantship as a primary means of support.

The most common methods of student support are either teaching or graduate assistantships. These jobs are awarded by the university in conjunction with the department in which the student is enrolled. Both provide compensation for the student in exchange for services to the university, but the responsibilities are quite different.

The teaching assistantship pays the student for efforts that are directly aimed at educating others. The graduate student will often teach an introductory class to a group of undergraduate students. Alternately, the TA might work as an aid to a professor, providing additional help to students, leading discussion or laboratory sessions, and doing

various tasks such as grading papers and proctoring exams. This experience can give the graduate student first hand experience in education. The process of planning a lecture or series of lectures, the ability to relate with less knowledgeable students, and the skill to teach them information in an effective manner can be a valuable asset for any hopeful academic. However, all of these efforts do take away time from the student's own research. According to the 1993 study, graduate students with teaching assistantships as the primary means of support took longer to complete the Ph.D. than students supported by research assistantships.

The research assistantship is often the more desirable means of support for a graduate student. The student's responsibilities revolve about actively participating in the research process. Some schools will allow laboratory rotations to count toward this requirement, whereas others assign a student to work as a technician in a lab. Upon deciding to devote dissertation efforts to a single lab, support from that researcher's grant will often replace the university funds, though this policy varies from school to school. The goal of a research assistantship is to give the student a concrete knowledge of some of the practical aspects of research as well as provide vitally needed funds. A research assistantship, unlike the teaching position, can often directly contribute to the student's own research, cutting down on the time needed to earn the Ph.D.

The compensation is generally the same for either type of assistantship. This compensation can take several forms. They often center around the student's tuition.

Tuition remission gives the student classes for free, tuition reduction gives the student a lower cost per credit and a tuition waiver grants the student a certain number of credits free of charge. This is a relatively inexpensive alternative for the university, since the space has already been allotted for classes, and the faculty paid for teaching them. Another type of compensation is the stipend, basically a sum of cash available to pay for classes (if they are not free), books, housing, bills, and other daily needs. In rare circumstances the graduate student's compensation will be treated as have a salaried position as an employee of the university. This option becomes more expensive for the university since there are often benefits that are given to staff members. A final form of reimbursement can be free or reduced cost use of university facilities such as free room and board. The form and magnitude of the compensation for the graduate student's work in an assistantship varies on the policies of the university and sometimes the department. If a school is trying to increase the size of a specific department, the financial benefits to students in that discipline could be more lucrative.

A student cannot affect the quality of a graduate program as directly as faculty or administrators, yet the student in part determines the school's quality. The student, more than anyone else wants the graduate experience to be worthwhile. After all, experiences in graduate school determine the course of the student's career, field of research, level of technical skill, and talent at science. Schools will be judged by the students that it graduates. The highly rated programs take less time to produce a doctorate than the lower rated programs (Goldberger, 1995). This fact is a reflection not only of the faculty and

the level of financial support, but of the dedication and effort of the student.

The final component of a multi-tiered plan of improvement is the program. The program is the point where university, faculty, and student come together and it is the unifying body under which all are judged by these studies of quality. Subjective indices from the reputational study examine the quality of the faculty, the effectiveness of the training, and the relative change in the program quality. Objective indices examine the number of faculty and students, percentage of full professors, number of PhD.s, percentage of minority students, and time necessary to earn the degree.

The administrative body of a program varies from school to school, some have committees, some have individual directors, some programs are subsidiaries of other departments, some are departments unto themselves. Whatever the configuration of the administrative body the responsibilities are the same. The administration is responsible for admitting students, determining the requirements for the Ph.D., overseeing the classes, coordinating logistical matters, and advising the university on topics such as faculty hiring, while all the time striving to improve the program. Each of these tasks takes the efforts of dedicated individuals who are willing to work within the political structure of the university and lead the program effectively.

The administration of a program has a wide array of options available to it that can increase quality in a manner that would be detectable in a study similar to those

already done. In order to change some of the objective statistics, the program can elect to hire new faculty, expand the number of students, or provide more financial aid. The reputational indices can also be affected by hiring well-known scholars to enhance the quality of the faculty and the research reputation of the program. The effectiveness of the program can also be improved by dedicating considerable time and effort to maximizing the opportunities and resources available to the students and hiring of faculty who are gifted educators.

Cartter's study examined the relationship between higher salaries and Highly rated programs. He found that the programs that offer more money can attract higher quality faculty (Diamond, 1989). This conclusion makes sense and the solution is an easy one, offer higher salaries. That action will attract faculty of greater quality and enhance program reputation. Of course, it is not that easy, money does not just appear. However, maintaining a better than average level of compensation should be a goal for any program on a crusade to improve itself.

In order to create the optimal educational environment, an internal analysis is often necessary, as well as a comparison to similar programs. Dr. Ronald Oppenheim of Wake Forest University, wished to identify shortcomings in his neuroscience program and examine how other schools administered their programs. He sent out a questionnaire to neuroscience program directors across the country inquiring about the curriculum, class offerings, and testing. Upon receipt and examination of these questionnaires, he

instituted a new examination at the end of the first year neuroscience classes (Oppenheim, 1997).

Hughes touched on a point that was not addressed in any subsequent studies, but is very important nevertheless. He believed that students should not be trained merely as researchers, but also as educators (Hughes, 1928). This idea is echoed in the Association of American Universities 1991 statement on improving graduate education. It states "Universities should

require all graduate students to teach, but should limit teaching time."{American Association of Universities, 1991). Though more and more PhD's are finding work outside academia, (Wilson, 1997). training in education an area that a program could devote some resources toward. Some students will receive teaching experience if they act as a TA, however, a required portion of the curriculum, in the form of workshops, classes, or teaching rotations, would enable all graduate students to learn the concepts of education.

In 1989, the Association of American Colleges began a three year project with the intention of educating students to educate as its goal. Graduate students took two seminars -- one on teaching, advising, and governance, and a second devoted to teaching subject matter in their disciplines. These seminars were coupled with visits to faculty of liberal arts colleges near the student's institution (Chronicle of Higher Education, 1989). More recently, some graduate schools have embraced this idea under a program called

"Preparing Future Faculty." Managed by the Association of American Colleges and Universities and the Council of Graduate Schools, this program provides seminars on education and projects to train graduate students as educators (Guernsey, 1998). One of the core components is the program's World Wide Web site which provides a forum for exchange of ideas, activities, program history and job resources. This page can be found at <<http://www.preparing-faculty.org>> (Association of American Colleges and Universities,1998). Both of these programs seek to give graduate students an idea of the responsibilities that they will have as educators and to provide them with the tools and the knowledge to teach effectively.

The computer and the use of online resources is rapidly becoming an essential part of the graduate experience. Though none of the indices examined computer resources, future studies should include an evaluation regarding them. The academic community is going online at a rapid rate, e-mail, web searches, and online journals are becoming part of the everyday tools used by a researcher. Though anybody with a copy of WordPerfect 8 or Microsoft Word 97 can create a web page, the servers in which to house that page, the Internet linkup, and the operating system are expensive investments best handled at the university level. E-mail alone is a wonderful tool for instantaneously sharing thoughts, charts, pictures, and data from anywhere in the world. New PhD's can search the internet for employment opportunities at sites such as the one created by the National Association of Graduate and Professional Students (Floyd, 1996). The World Wide Web can allow access to hundreds of resources at the click of a mouse. The federal

government has committed $50 million in grants to universities that are developing "digital libraries," that is, libraries whose resources are available over computer networks of the internet (Kiernan, 1998). All of these wondrous electronic advances can be used by the student as tools to expand the graduate experience.

However, along with the ease of information acquisition that technology offers, students must be aware of how to examine that information in a critical fashion. Unregulated sites could be filled with misleading or even false information (McBride, 1998). A workshop or even an entire class devoted to the analysis of the usefulness of information resources in scholarly pursuits would be a valuable addition to a neuroscience curriculum. Rapid access to current knowledge is a must for today's scientist or graduate student, and much like a well-maintained library, with a wide variety of current publications, a computer networked with the Internet is an investment that a university can make to expedite the flow of information from the world to the student.

In the spirit of wide access to research, Virginia Tech is requiring graduate students to submit their thesis digitally. Some students have grumbled about the difficulty of learning HTML or lack of available equipment, but the university has provided training and software to ease those burdens. The use of computer technology has allowed students to present work with color pictures, website links, sound clips, and even videos. Dissertations that are posted on the web are also easily accessible to other researchers, all that is necessary is an Internet link (Young, 1998). This example of

technological innovation is just a precursor of what science will be like in the future when research and information are only as far away as a computer mouse. Students today should learn how to use technological research tools as a part of graduate education, just as they learn lab procedures and grant writing.

Working together, the university, the faculty, the student and the program can all contribute to an increase in quality of a graduate program. Studies by Goldberger and Maher and all of their predecessors show what points are used to define quality and forecast what future measurements will be taken. Beyond the current focus on the indices that are explored in these historical studies, a neuroscience program must look toward the future and plan to be judged on its computer resources, financial structure, and the fate of its graduates. A program must do more than simply react to the standards that have been set, it must predict what is on the horizon and adapt to new ideas. This forward thinking, coupled with the determination to make the effort to change, while guided by the ideas of the past will enhance the quality of any neuroscience program with a supportive university, a dedicated faculty, and enthusiastic students.

Appendix

Carnegie Classifications

This description is found at a World Wide Web site at the following address:

http://webserv.educom.edu/members/CarnegieCls.html

It has been edited for space and reformatted from HTML.

Carnegie Classification

The Carnegie Foundation for the Advancement of Teaching produces a "Carnegie Classification" list which groups accredited institutions of higher education into 11 categories, based largely on their academic missions.

Carnegie's classification is a key resource for academe. Institutions are classified according to the highest level of degree they award, the number of degrees conferred by discipline, and, in some cases, the amount of federal research support they receive and the selectivity of their admissions.

Research Universities I

These institutions offer a full range of baccalaureate programs, are committed to graduate education through the doctorate, and give high priority to research. They award 50 or more doctoral degrees (1) each year. In addition, they receive annually $40-million or more in federal support.(2)

Research Universities II

These institutions offer a full range of baccalaureate programs, are committed to graduate programs through the doctorate, and give high priority to research. They award 50 or more doctoral degrees (1) each year. In addition, they receive annually between $15.5

million and $40-million in federal support.(2)

Doctoral Universities I

In addition to offering a full range of baccalaureate programs, the mission of these institutions includes a commitment to graduate education through the doctorate. They award at least 40 doctoral degrees(1) annually in five or more disciplines.(3)

Doctoral Universities II

In addition to offering a full range of baccalaureate programs, the mission of these institutions includes a commitment to graduate education through the doctorate. They award at least 10 doctoral degrees--in three or more disciplines(3) --or 20 or more doctoral degrees in one or more disciplines.

Master's Degree (Comprehensive) Universities and Colleges I

These institutions offer a full range of baccalaureate programs and are committed to graduate education through the master's degree. They award 40 or more master's degrees annually in three or more disciplines.(3)

Master's Degree (Comprehensive) Universities and Colleges II

These institutions offer a full range of baccalaureate programs and are committed to graduate education through the master's degree. They award 20 or more master's degrees annually in one or more disciplines.(3)

Baccalaureate (Liberal Arts) Colleges I

These institutions are primarily undergraduate colleges with major emphasis on baccalaureate degree programs. They are selective in admissions and award 40 per cent or more of their baccalaureate degrees in liberal arts fields.(4)

Baccalaureate (Liberal Arts) Colleges II

These institutions are primarily undergraduate colleges with major emphasis on baccalaureate degree programs. They are less selective in admissions or they award less than 40 per cent of their baccalaureate degrees in liberal arts fields.(4)

Associate of Arts Colleges

These institutions offer associate of arts certificate or degree programs and, with few exceptions, offer no baccalaureate degrees.(3)

Professional Schools and Specialized Institutions

These institutions offer degrees ranging from the bachelor's to the doctorate. At least 50 per cent of the degrees awarded by these institutions are in a specialized field.

Specialized institutions include:

Theological seminaries

This category includes institutions at which the primary purpose is to offer religious instruction or train members of the clergy.

Medical Schools and Medical Centers

These institutions award most of their professional degrees in medicine. In some instances, their programs include other health professional schools, such as dentistry, pharmacy, or nursing.

Other Separate Health Profession Schools

Institutions in this category award most of their degrees in such fields as chiropractic, nursing, pharmacy, or podiatry.

Schools of Engineering & Technology

The institutions in this category award at least a bachelor's degree in programs limited almost exclusively to technical fields of study.

Other Specialized Institutions

Institutions in this category include graduate centers, maritime academies, military institute, and institutions that do not fit any other classification category.

Schools of Business & Management

The schools in this category award most of their bachelor's or graduate degrees in business or business-related programs.

Other/Non-Specified Institutions

Schools of art, music, and design: Institutions in this category award most of their bachelor's or graduate degrees in art, music, design, architecture, or some combination of such fields.

Schools of law: The schools in this category award most of their degrees in law. The list includes only institutions that are listed as separate campuses in the 1994 Higher Education Directory.

Teachers colleges: Institutions in this category award most of their bachelor's or graduate degrees in education or education-related fields.

Tribal colleges: These colleges are, with few exceptions, tribally controlled and located on reservations. They are all members of the American Indian Higher Education Consortium.

Notes:

(1) Doctoral degrees include Doctor of Education, Doctor of Juridical Science, Doctor of Public Health, and the Ph.D. in any field.

(2) Total federal obligation figures are available from the National Science Foundation's annual report called "Federal Support of Universities, Colleges and Nonprofit Institutions." The years used in averaging total federal obligations are 1989, 1990, and

1991.

(3) Distinct disciplines are determined by the U.S. Department of Education's "Classification of Instructional Program's" 4-digit series.

(4) The liberal-arts disciplines include area and ethnic studies, English language and literature, foreign languages, letters, liberal and general studies, life sciences, mathematics, multi- and interdisciplinary studies, philosophy and religions, physical sciences, psychology, social sciences, and the visual and performing arts. The occupational and technical disciplines include agriculture, allied health, architecture, business and management, communications, conservation and natural resources, education, engineering, health sciences, home economics, law and legal studies, library and archival sciences, marketing and distribution, military sciences, protective services, public administration and services and theology.

(5) This group includes community, junior, and technical colleges.

References

Association of American Colleges and Universities (1998). *Preparing Future Faculty* [Web site]. Washington, D.C.

Association of American Universities (1991). AAU Policy on Improving Graduate Education. *Chronicle of Higher Education, XXXVII* (January 23, 1991).

Association of Neuroscience Departments and Programs.(1994). *Neuroscience Training Programs in North America.* (1994 ed.). Washington, D.C.: Association of Neuroscience Departments and Programs.

Barrons. (1996). *Barron's Guide to Law Schools.* (12 ed.). Hauppauge, NY: Barron's Educational Series, Inc.

Basinger, J. (1998). *Student Loan Debt Burdens More Than Half of All Graduates, Government Report Says* [Website].

Carnegie Institute (1997). *Carnegie Classifications* [Web Site]: Carnegie Foundation for the Advancement of Teaching.

Cartter, A. (1964). *An Assessment of Quality in Graduate Education: A Comparitive study of Graduate Departments in 29 Academic Disciplines*: American Council on Education.

Chronicle of Higher Education. (1989). Program Teaches Graduate Students About Professorate. *Chronicle of Higher Education, XXXIII* (October 18, 1989), A19.

Cordes, C. (1997). Science Board Considers Federal Role in Improving Graduate Education. *Chronicle of Higher Education, XLIV*(October 24, 1997), A27.

Diamond, R. M. (1989). *Designing and Improving Courses and Curricula in Higher Education.* (1 ed.). San Francisco, California: Jossey-Bass Inc.

Dopp, S. (1997). Administrative Assistant, Acquisitions Dept: University of Maryland. Personal Communication.

Floyd, B. (1996). Graduate-Student Organization Creates Electronic Job Bank. *Chronicle of Higher Education, XLII* (February 2, 1996), A22.

Goldberger, M. and Maher B. (1995). *Research-Doctorate Programs in the United States: Continuity and Change.* Washington, D.C.: National Research Council.

Gose, B. (1997). Average Cost of Tuition is Up 5% for 1997-1998. *Chronicle of Higher Education, XLIV* (October 3, 1997), A49.

Gourman, J. (1993). *The Gourman Report--A Rating of Graduate and Professional Programs in American & International Universities*. (6 ed.). Los Angeles: National Education Standards.

Guernsey. (1998). Web Site Helps Ph.D. Students Prepare for Faculty Life. *Chronicle of Higher Education, XLIV*(19), A23.

Hughes, R. (1928). Report of the Committee on Graduate Instruction. *American Universities and Colleges*. Washington, D.C.: American Council on Education.

Hughes, R. (1934). Report of the Committee on Graduate Instruction. *The Educational Record, 15*(2), 192-234.

Jacobsen, D., Eggen, P., and Kauchak, D. (1989). *Methods for Teaching--A Skills Approach*. (3 ed.). New York City: Macmillan Publishing Company.

Jones, L., Lindzey, G. and Coggeshall, P. (1982). *An Assessment of Research-Doctorate Programs in the United States: Biological Sciences*. (Vol. 4). Washington, D.C.: National Academy Press.

Keniston, H. (1959). *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*. New York, New York: American Book-Stratford Press, Inc.

Kiernan, V. (1998). U.S. Plans to Offer $50-Million in Grants to Support Research on Digital Libraries. *Chronicle of Higher Education, XLIV* (February 27, 1998), A27.

King, R., Paget, N., and Ingvarson, L. (1993). An Interdisciplinary Course Unit in Basic Pharmacology and Neuroscience. *Medical Education, 27*, 229-237.

Mancall, E., Murray, T., Swick, H., Miller, J., Smith, D., Weiss, M. (1987). A Model Clinical Neuroscience Curriculum. *Neurology, 37* (November), 1697-1699.

McBride, K. and Dickstein, R. (1998). The Web Demands Critical Thinking by Students. *Chronicle of Higher Education, XLIV* (28), B6.

McNeil, J. (1990). *Curriculum--A Comprehensive Introduction*. (4 ed.). Los Angeles, California: Harper Collins Publishers.

Oppenheim, R. (1997). Director, Neuroscience Program, Bowman-Gray Medical School, Wake Forest University. Personal Communication.

Presti, D. and Barrie, J. (1996). The World Wide Web as an Instructional Tool. *Science, 274* (October 18), 371-372.

Roose, K. and Andersen, C. (1970). *A Rating of Graduate Programs*: American Council on Education.

Rosovsky, H. (1990). *The University An Owner's Manual*. (1 ed.). New York City: W.W. Norton and Company, Inc.

Schwartz, S. and Freidman, M. (1992). *A Guide to NIH Grant Programs*. New York: Oxford University Press.

Shapiro, J. (1997). University Libraries: The 7-per-cent Solution. *Chronicle of Higher Education, XLIV* (December 12, 1997), B4.

Smith, D. O. (1993). Gender Disparity in the Academic Pipeline: Women in Neuroscience. *Synapse, 14*, 332-334.

Verma, A. (1996). Evaluation of USUHS Neuroscience Training Program .

Weinstein, G., Hardin, J., & Weinstein, M. (1976). *Education of the Self*. Amherst, MA: Mandala.

Wiles, J., & Bondi, J. (1993). *Curriculum Development--A Guide to Practice*. (4 ed.). New York City: Macmillan Publishing Company.

Wilson, R. (1997). Career Success for Some Ph.D.'s Comes by Leaving Academe Behind. *Chronicle of Higher Education, XLIV* (October 31, 1997). A54.

Young, J. (1998). Requiring Thesis in Digital Form: The First Year at Virginia Tech. *Chronicle of Higher Education, XLIV* (February 13, 1998), A29.